# Implementation of Object Boundary Measurement Method Using Image Processing of Multiple Line Lasers

Janghoon Jeong[1], Dong-Hwan Kim[2], and Seong-Ho Son[1,2,*]
[1]*Department of ICT Convergence, Soonchunhyang University, Asan, Republic of Korea*
[2]*Department of Mechanical Engineering, Soonchunhyang University, Asan, Republic of Korea*
*Contact: son@sch.ac.kr

*Abstract*— **This paper introduces a novel approach for real-time body boundary estimation in microwave imaging using line laser technology. Our method utilizes line lasers and cameras in an experimental setup involving a water-filled tank to simulate the presence of a human body. The collected images are processed through binarization, thinning, and calibration to estimate the body's boundaries. In cases of data insufficiency, interpolation techniques fill in the gaps. Experimental results demonstrate the effectiveness of our approach, achieving a precise boundary estimation with a root mean square error (RMSE) of 1.27 mm and a similarity rate of 98.97%. This innovative approach offers potential advancements in the field of cancer treatment by facilitating effective and precise tumor targeting.**

## I. Introduction

Phased array technology, traditionally for communication and radar, now used in cancer treatment with microwaves. Typical system uses 100 MHz frequency, while Focused Microwave Thermotherapy (FMT) employs 900 MHz for precise tumor targeting [1–4]. Patient shape and tumor location data are crucial for accurate treatment.

Previous studies [1–3,5] and phantom experiments [2,4] confirm the need for patient information, though hard to obtain in real scenarios. FMT system incorporates a water bolus for skin cooling and electromagnetic matching [6], making external body shape estimation challenging. Microwave tomography struggles with boundary reconstruction due to its inverse nature [3]. Ultrasound-microwave fusion is complex [7] requiring two modalities. Radar-based studies aim to estimate boundaries through experiments and analysis [8], facing sensor constraints and estimation errors. Point laser systems [9] are slow (up to 30 min), limiting real-time tracking of body boundaries. Inefficient tumor treatment is a concern [10].

This paper introduces a line laser vision method to measure body boundaries in real-time with advantages like capturing multiple points simultaneously. Challenges include correcting distorted camera images due to water bolus refraction. The laser can't penetrate the bolus, and data insufficiency and body movement are addressed.

## II. Materials and Methods

### A. Estimation Configuration

The experimental configuration for proof-of-concept is shown in Figure 1. The configuration primarily includes line

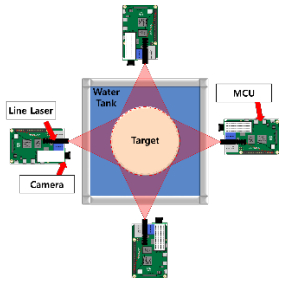



Fig. 1 An experimental configuration for the proposed body boundary estimation

sides of the body. To manage these devices and transmit the camera images to the main computer, a microcontroller is employed.

A pair of line lasers and cameras is affixed to the exterior of the tank to illuminate the body's cross-section. The tank is filled with water to simulate a mass of water around the body. The camera image is wirelessly transmitted to the main computer, which then executes an algorithm to extract the body's boundary.

### B. Process of Boundary estimation

The line laser area is extracted from the collected images, and the body boundaries are estimated through binarization, thinning, and calibration. By merging data from all locations that have gone through this process, we can estimate the boundaries of the entire human body.

Additionally, in case of insufficient data, the boundaries of the human body are estimated by filling in the missing parts through interpolation. Figure 2 shows the flow chart where the boundaries are estimated.

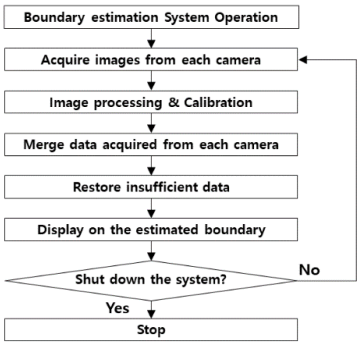



Fig. 2 Flowchart of Experimental Subject Boundary Estimation

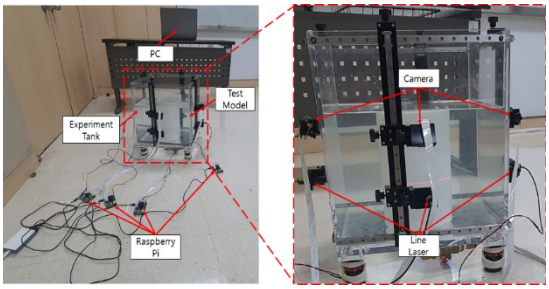



Fig. 3 Experimental Setup for Boundary Estimation System Application

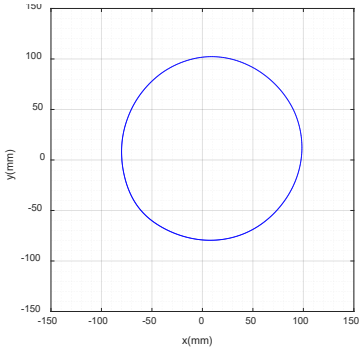



Fig. 4 Result of estimating a cylinder with a diameter of 180mm

### C. Performance Indicators

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i - \hat{r}_i)^2} \quad (1)$$

$$Similarity\ rate = \left(1 - \frac{1}{N}\sum_{i=1}^{N}\left|\frac{r_i - \hat{r}_i}{r_i}\right|\right) \times 100 \quad (2)$$

where $r_i$ and $\hat{r}_i$ are the distances calculated in polar coordinates of the actual boundary and estimated boundary, respectively.

## III. Result and Discussion

To verify the proposed method, an experiment was conducted after filling the experimental tank with water and placing a cylinder with a diameter of 180 mm. Fig. 3 shows the experimental setup for verification and Fig. 4 shows the boundary of the estimated cylinder. The effectiveness of the estimated results is calculated using Eq. (1) and Eq. (2). The calculated results showed that precise boundary estimation was possible with an RMSE of 1.27 mm and a Similarity Rate of 98.97%.

## IV. Conclusion

In this work, we proposed and verified a new method for estimating patient body boundaries during cancer treatment using microwaves. This method aims to eastimate body boundaries in real time using line laser vision technology, and this method has several advantages. The experimental results showed that the proposed method was capable of precise boundary estimation, with a root mean square error (RMSE) of 1.27 mm and a similarity rate of 98.97%. This is expected to help accurately measure the patient's body boundaries and effectively perform microwave therapy.

## Acknowledgment

This work was supported by Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) (P0012724, The Competency Development Program for Industry Specialist.

## References

[1] J. Y. Kim, S-I. Kim, K.-J. Lee, B.-R. Kim, N. Simonov, J. S. Yoon, N. Kim, and S.-H. Son "Computational study of focused microwave thermotherapy for knee pathological treatment", *IET Microw. Antennas. and Propag*, vol. 12, pp. 1901–1907, 2018.

[2] J.-Y. Kim, K.-J. Lee, B.-R. Kim, S.-I. Jeon and S.-H. Son, "Numerical and experimental assessments of focused microwave thermotherapy system at 925 MHz", *ETRI J.*, vol. 41, no. 6, pp. 850-862, 2019.

[3] N. Simonov and S.-H Son, "Focused microwave thermotherapy technique based on microwave tomographic imaging," *Electron. Lett.*, vol. 55, No. 11, pp.633–634, 2019.

[4] K.-J. Lee, J.-Y. Kim and S.-H. Son, "Experimental phantom test of 925 MHz microwave energy focusing for non-invasive local thermotherapy," *Results in Physics*, vol. 59, No. 105585, 2022.

[5] E. Zastrow, S. Hagness and B. VanVeen, "3D computational study of non-invasive patient-specific microwave hyperthermia treatment of breast cancer", *Phys. Med. Biol.*, vol. 55, pp. 3611-3629, 2010.

[6] M. A. Ebrahimi-Ganjeh and A. R. Attari, "Study of water bolus effect on SAR penetration depth and effective field size for local hyperthermia", *Prog. Electromagn. Res. B*, vol. 4, pp. 273-283, 2008.

[7] M. Omer, P. Mojabi, D. Kurrant, J. LoVetri and E. Fear, "Proof-of-concept of the incorporation of ultrasound-derived structural information into microwave radar imaging", *IEEE J. Multiscale Multiphys. Comput. Tech.*, vol. 3, pp. 129-139, Aug. 2018.

[8] D. Winters, J. Shea, E. Madsen, G. Frank, B. Van Veen and S. Hagness, "Estimating the breast surface using UWB microwave monostatic backscatter measurements", *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 247-256, Jan. 2008.

[9] T. Cameron, T. C. Williams, J. Bourqui, M. Okoniewski, and E. C. Fear, "Comparison of microwave and laser surface detection for microwave imaging system", *Proc. 13th Int. Symp. Antenna Tech. Applied Electromag.*, pp. 1-4, 2009.

[10] Q.-S. Chen, M. S. Weinhous, F. C. Deibel, J. P. Ciezki and R. M. Macklis, "Fluoroscopic study of tumor motion due to breathing: Facilitating precise radiation therapy for lung cancer patients", *Med. Phys.*, vol. 28, no. 9, pp. 1850-1856, 2015.

# A Conceptual Design of Wilkinson Divider/Balun for Dual Functions

Hyungzun Mun[1], Jeongho Park[2], Sang-Min Han[2], Dal Ahn[1,2] and Jongsik Lim[1,2*]
[1]*Dept. of Electrical Comm. System Engineering, Soonchunhyang Univ., Asan, Chungnam, 31538, Rep. Of KOREA*
[2]*Dept. of ICT Convergence, Soonchunhyang Univ., Asan, Chungnam, 31538, Rep. Of KOREA*
**Contact: jslim@sch.ac.kr, phone +82-41-530-1332*

***Abstract* – This paper presents the conceptual design of a Wilkinson divider which have additional balun functions. The well-known structure of Wilkinson power divider is combined with SPDT (single pole double throw) switch and a phase out of phase converting structure. Dual functions as a power divider and balun are realized by switching the path of output signal to the proper port using the single pole double throw switch. The conceptual design of the proposed divider/balun is well verified with the required power division ratio, port matching at all ports, isolation between output ports, and in-phase/out-of-phase characteristics according to the switching status of the SPDT switch.**

***Keywords* – Wilkinson divider, balun, RF switch**

## I. Introduction

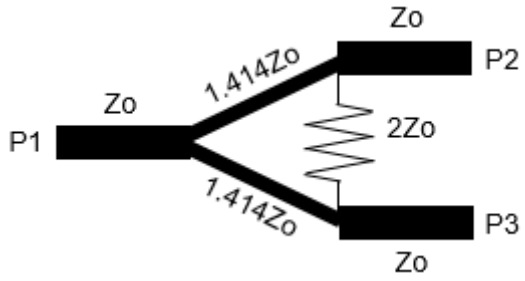
In the system of various wireless communication systems, microwave Wilkinson power divider is one of the most widely used wireless circuits [1,2]. When the phase difference between the two power-divided outputs is 180°, it serves as a microwave balun circuit [3]. If a 180° phase converting structure is inserted into one of the output ports of the Wilkinson divider, this means the power divider operates as a balun. In this work, the basic concept of the dual functions of the Wilkinson divider/balun is discussed by combining the Wilkinson divider, phase converting structure and SPDT (single pole double throw) switch.

## II. Microwave Wilkinson Power Dividers

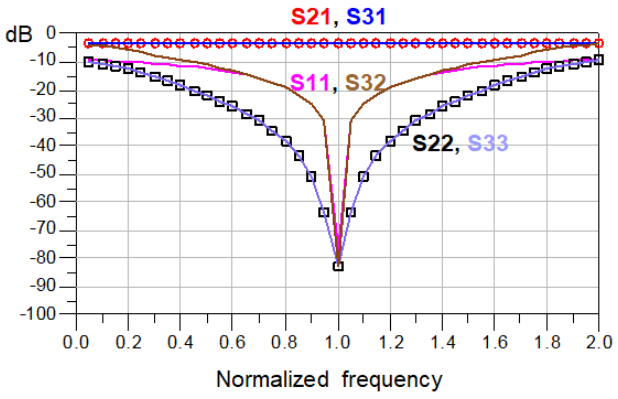
Fig. 1(a) illustrates the basic topology of microwave Wilkinson power dividers. When the input signal is applied to port1, the signal is divided to the port2 and port3. Since this is a simple power division, there is no phase difference between the two outputs [2,4]. Fig. 1(b) shows typical S-parameters of Wilkinson power dividers, and Fig 1(c) the phase difference between output ports.

## III. Microwave Balun with Wilkinson Divider Structure

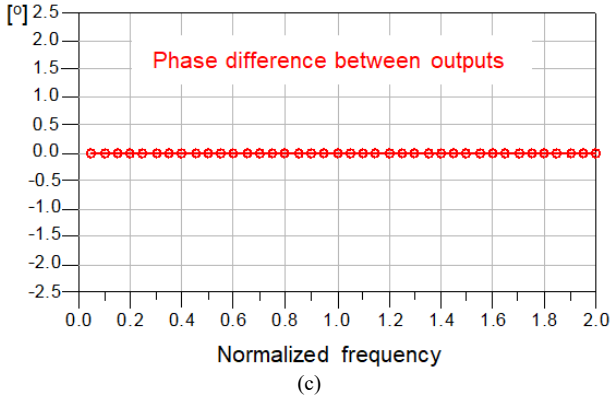
Fig. 2(a) is the basic circuit of Wilkinson balun. When the phase difference between outputs ports of Wilkinson divider is 180°, it operates as balun circuit [3]. Fig. 2(b) shows the typical S-parameters of Wilkinson baluns, and Fig. 2(c) the phase difference between output signals. It is noted that the S-parameters of the balun in Fig. 2(b) are exactly the same as Fig. 1(b). However, in Fig. 2(c), it is shown that the Wilkinson balun has opposite phase characteristics to Wilkinson dividers illustrated in Fig. 1(c).




(a)

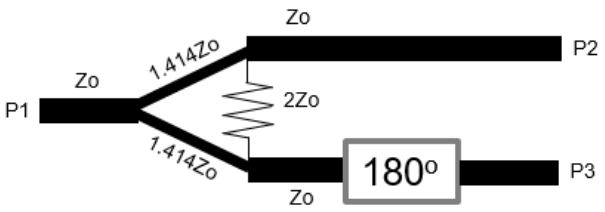



(b)

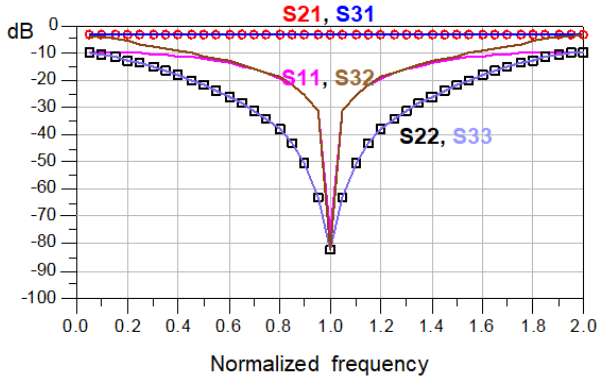



(c)

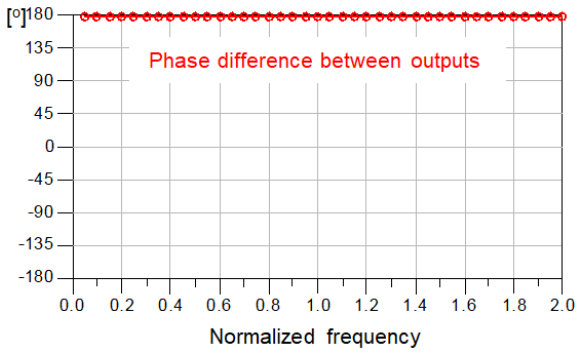
Fig. 1 Normal Wilkinson power dividers (a)basic schematic (b)ideal performances (c)in-phase characteristics between output ports




(a)




(b)




(c)

Fig. 2 Normal Wilkinson balun circuit (a)basic schematic (b)ideal performances (c)out-of-phase characteristics between output ports

## IV. Conceptual Design of Wilkinson Divider/Balun

One can compare the divider in Fig. 1(a) and balun in Fig. 2(a) easily. The Wilkinson structure operates as a power divider or balun whether there exists a 180° phase converting structure on one of output paths or not. Therefore, if the phase converting structure is adopted selectively, the Wilkinson structure may be used as a power divider or balun with dual functions.

The switching of dual functions is obtained conceptually from RF switches. Fig. 3 shows the conceptual block diagram of SPDT (single pole double throw) switches. The input signal is transmitted to output1 or output2 by selecting the switching function.

Fig. 4 is the conceptual block diagram of 0°/180° phase controlling structure. The 0° means there is no specific device which cause a phase change.

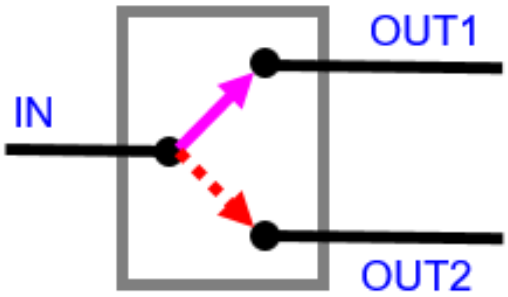



Fig. 3 Basic functions of RF SPDT switch

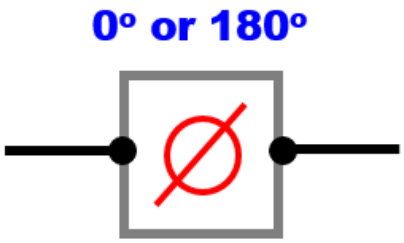



Fig. 4 Concept of the in-/out-of-phase controller

Fig. 5 is the proposed schematic of the Wilkinson divider/balun, which shows the conceptual dual functions of the Wilkinson power divider and balun. An SPDT switch is connected to one of the output ports, and a 180° phase converting structure is attached to the path of OUT2 of the switch. If the switch is connected to OUT1, the two output ports (P2, P3D) of the Wilkinson divider are in-phase, and it operates as a power divider just like the circuit in Fig. 1(a). To the contrary, when the switching function goes to OUT2, the two output ports (P2, P3B) are in out-of-phase. Now it works as a balun, which is so similar to Fig 2(a).

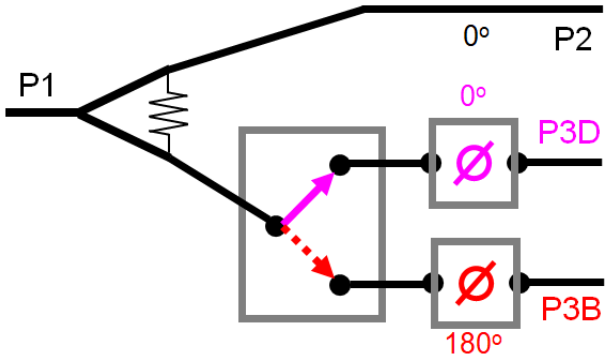
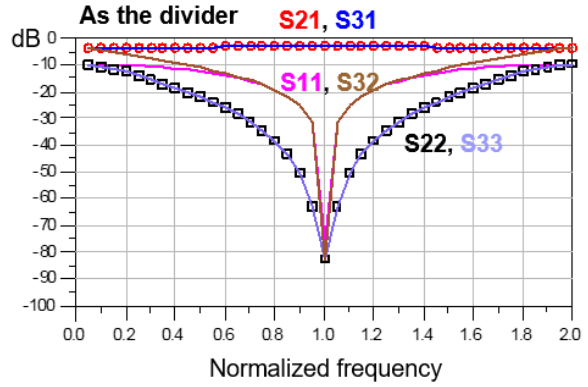
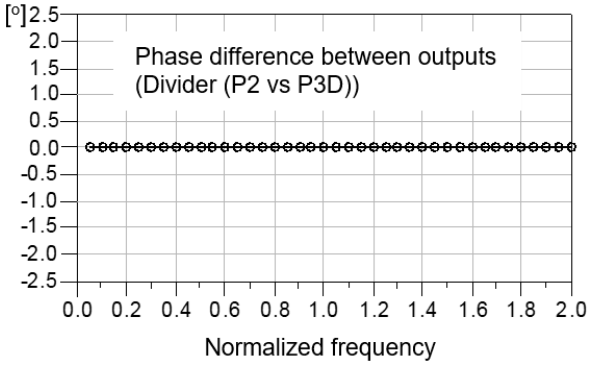


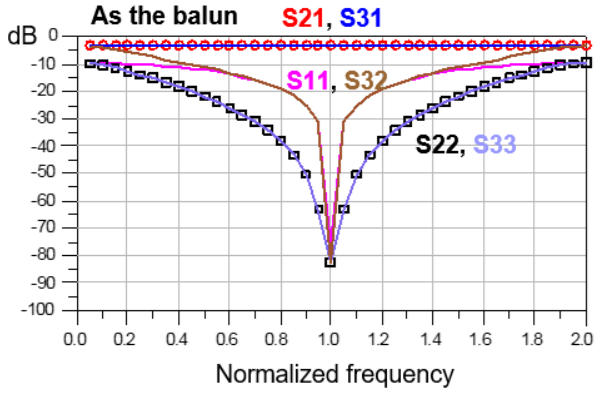

Fig. 5 Proposed concept of the Wilkinson divider/balun having dual functions



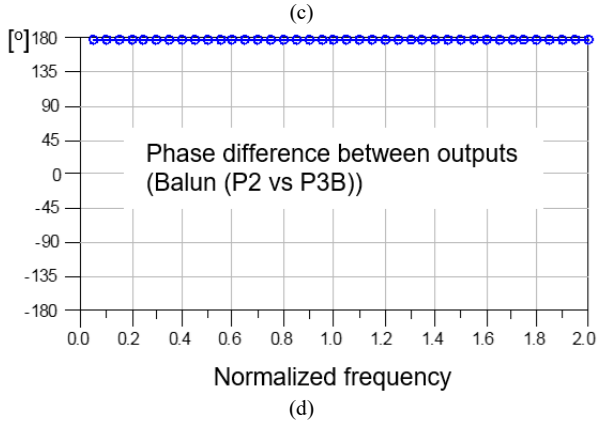

(a)




(b)




(c)




(d)

Fig. 6 Dual functions of the conceptually proposed Wilkinson divider/balun (a)power division as the divider (b)in-phase characteristics as the power divider (c)power division as the balun (b)out-of-phase characteristics as the balun

Fig. 6 illustrates the S-parameters and phase difference between output ports according to the switching status. Firstly, Fig. 6(a) and Fig. 6(b) show the performances at P2 and P3D as a power divider. The graphs exhibit the same power division and in-phase characteristics as demonstrated in the Fig. 1. Fig. 6(c) and Fig. 6(d) also show the power division and out-of-phase between P2 and P3B when operating as a balun. These graphs illustrate the same power division and out-of-phase characteristics as discussed in the Fig. 2.

## V. Conclusions

In this paper, a conceptual circuit that has dual functions as a power divider and a balun by combining an RF switch and a 180° phase converting structure for Wilkinson power dividers. The proposed circuit has dual functions as a power divider and a balun by selecting the switching functions of a SPDT switch. The authors are going to study further in order to implement the proposed conceptual circuit practically and verify its electrical performances. It is expected the dual function as a power divider and a balun would be obtained by adopting and a proper RF SPDT switch and phase converting structure.

## Acknowledgment

This work was supported by ICAN (ICT Challenge and Advanced Network of HRD) program supervised by the IITP(IITP-2023-2020-0-01832) funded by the Ministry of Science and ICT (MSIT), and Regional Innovation Strategy (RIS) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE)(2021RIS-004).

## References

[1] D. M. Pozar, *Microwave Engineering (4/e)*, Ch.7, John Wiley and Sons, Inc., New York, 2011.

[2] E. J. Wilkinson,"An N-way hybrid power divider," *IEEE Trans. Mi-crow. Theory Tech*., vol. MTT-8, no. 1, pp. 116–118, Jan. 1960.

[3] J. Lim, U. Park, Y. Jeong, K. Choi, D. Ahn, S. Oh, and J. Koo, "800-5000MHz ultra-wideband CPW balun," *IEE Electronics Letters*, vol. 42, no.18, pp. 1037-1039, Aug. 2006.

[4] S. Horst, R. Bairavasubramanian, M. Tentzeris and J. Papapolymerou, "Modified Wilkinson Power Dividers forMillimeter-Wave Integrated Circuits," *IEEE MTT Trans. Microwave Theory Tech*., Vol. 55, no. 11 pp. 2439–2446, Nov. 2007.

# Quasi Convex Adaptive Sliding Mode Control With Time-Varying Disturbances

Jin Woong Lee[1] and Seok Young Lee[1,*]
[1]*Department of ICT Convergence Engineering, Soonchunhyang University, Asan, South Korea*
*Contact: suk122@sch.ac.kr

*Abstract*— **In this paper, a new quasi convex function based adaptive control algorithm is proposed for robot manipulator. This controller This controller reduces the overestimation of the control gain and keeps the sliding surface to arbitrarily small vicinity of the zero.**

## I. INTRODUCTION

Many robot manipulators have been researched and utilized in a wide range of fields, including IoT and industrial environments. However, the control performance of robot manipulators is affected by undesired factors such as nonlinear friction, unknown external disturbances, mechanical noises, and system uncertainties. To deal with the aforementioned factors, a nonlinear robust control algorithm called sliding mode control (SMC) was proposed [1]. The SMC guides the state of the system to move along the sliding surface, keeps the system to a goal state, and maintains stable motion. However, since the SMC requires the knowledge of an upper bound of uncertainties and disturbances, the SMC utilization in a real environment is a difficult task.

So, the adaptive sliding mode control (ASMC), which combines the SMC and an adaptive gain, was proposed [2]. Because the ASMC dynamically modifies the gain, the ASMC does not require the knowledge of an upper bound of uncertainties in advance [3]. The ASMC may cause overestimation. And this increases chattering and reduces control performance.

To reduce overestimation, a function-based adaptive gain control such as a barrier function adaptive sliding mode control (BFASMC) was proposed [4]. This approach ensures that the sliding surface converges to zero without significantly overestimating the gain. In a function-based adaptive gain, the function design significantly affects control performance. This paper proposes a new function-based adaptive sliding mode control with a quasi-convex function [5].

The quasi-convex adaptive sliding mode control (QCASMC) adjusts the adaptive gain according to the quasi-convex function when the sliding surface is near zero and this reduces overestimation. This paper compares the BFASMC and the QCASMC on a 2-DOF robot manipulator.

The rest of this paper is organized as follows. Section II describes manipulator system and ASMC. Section III designs the QCASMC and proves the stability of QCASMC. Section IV simulates and compares QCASMC and BFASMC on 2-DOF manipulator. Finally, Section V concludes this paper.

## II. PRELIMINARIES

### A. Manipulator system

Consider the Euler-Lagrange dynamic equation of the n-degree of freedom robot manipulators as

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) + F(\dot{q}) = \tau, \quad (1)$$

where system parameters $q, \dot{q}, \ddot{q} \in \mathbb{R}^n$ are the angular position, velocity, and acceleration vectors of each manipulator joint, correspondingly. $M(q) \in \mathbb{R}^{n\times n}$ is the inertia matrix of the robot manipulator, $C(q,\dot{q}) \in \mathbb{R}^{n\times 1}$ is the centripetal Coriolis matrix, $G(q) \in \mathbb{R}^{n\times 1}$ is the gravitational vector, $F(\dot{q}) \in \mathbb{R}^{n\times 1}$is the friction force vector. $\tau \in \mathbb{R}^{n\times 1}$ is the control input torque of each manipulator joint. Multiplying both sides of (1) by $M^{-1}$ and rewriting as $\ddot{q}$, we have

$$\ddot{q} = M^{-1}(\tau - C(q,\dot{q}) - G(q) - F(\dot{q})). \quad (2)$$

### B. Adaptive sliding mode control

In this paper, we define system error as $e = q_d - q$, where $q_d \in \mathbb{R}^{n\times 1}$ is the desired angular position. The sliding surface is chosen as

$$s = \dot{e} + \Lambda e, \quad (3)$$

where $\Lambda \in \mathbb{R}^{n\times n}$ denotes the nonzero positive diagonal matrix. Using the system dynamics (2), the time derivative of the sliding surface is

$$\begin{aligned} \dot{s} &= \ddot{e} + \Lambda\,\dot{e} = \ddot{q}_d - \ddot{q} + \Lambda\,\dot{e}, \\ &= \ddot{q}_d - M^{-1}(\tau - C - G) + \Lambda\,\dot{e}. \end{aligned} \quad (4)$$

From the dynamics (2), we construct the following control:

$$\begin{aligned} \tau &= \tau_0 + \tau_1, \\ \tau_0 &= C + G + M(\ddot{q}_d + \Lambda\,\dot{e}), \\ \tau_1 &= M\hat{K}\,sgn(s), \end{aligned} \quad (5)$$

where the control gain $\hat{K} \in \mathbb{R}^{n\times n}$ is a positive diagonal switching gain matrix. The $sgn(s) \in \mathbb{R}^{n\times 1}$ is defined by

$$sgn(s) = \begin{cases} 1, & s \geq 0 \\ -1, & s < 0. \end{cases} \quad (6)$$

## III. Main result

### A. Design QCASMC

The control gain $\hat{K}$is controlled by a new adaptive law. The new adaptive law, including the QC, is defined as follows:

$$\dot{\hat{K}} = \begin{cases} K_a, \dot{\hat{K}_a} = \alpha|s|, & |s| \geq \epsilon \\ K_b = 1 - e^{-\beta s^2}, & |s| < \epsilon. \end{cases} \quad (7)$$

The $\epsilon$ is small positive number near zero, the α is tunable positive gain for adaption speed, and the β is tunable gain of slope of the convex part of the QC. The proposed new adaptive law reduces the adaptive gain K^to zero when s is near zero.

### B. Stability analysis

The stability of the QCASMC is verified via the Lyapunov stability theorem. According to [3], the sliding surface will reach $\frac{\epsilon}{2}$ in finite time. Considering $|s| < \epsilon$, the Lyapunov function is defined as follows:

$$V = \frac{1}{2}s(t)^T s(t) + \frac{1}{2}(K_b(s(t)) - K_b(s(0))^2. \quad (8)$$

The derivate of Lyapunov function is given as follows

$$\begin{aligned} \dot{V} &= \dot{s}s + (K_b - K_b(s(0))\dot{K_b} \\ &= s\big(-K_b sgn(s)\big) + (K_b(s) - K_b(s(0))\dot{K_b} \\ &= -K_b|s| - K_b(\beta s^2 e^{-\beta s^2} K_b|s|) \\ &\leq 0. \end{aligned} \quad (9)$$

Thus, we prove that the stability of the QCASMC and gain of the QCASMC is optimized to zero when $|s| < \epsilon$.

## IV. Simulation result

We compare the BFASMC and the QCASMC on a 2-DOF robot manipulator. For the comparison, the 2-DOF robot manipulator is given by [6]

$$M(q) = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}, C(q,\dot{q})\dot{q} = \begin{bmatrix} C_{11} \\ C_{21} \end{bmatrix}, \quad G(q) = \begin{bmatrix} G_{11} \\ G_{21} \end{bmatrix}, F(\dot{q}) = \begin{bmatrix} F_{12} \\ F_{22} \end{bmatrix}, \quad (10)$$

with

$$\begin{aligned} M_{11} &= {l_2}^2 m_2 + 2l_1 l_2 m_2 \cos(q_2) + {l_1}^2(m_1 + m_2), \\ M_{12} &= {l_2}^2 m_2 + l_1 l_2 m_2 \cos(q_2), \\ M_{21} &= M_{12},\ M_{22} = {l_2}^2 m_2, \\ C_{11} &= -l_1 l_2 m_2 \sin(q_2)\, {\dot{q}_2}^2 - 2l_1 l_2 m_2 \sin(q_2)\, {\dot{q}_1}^2 {\dot{q}_2}^2, \\ C_{21} &= l_1 l_2 m_2 \sin(q_2)\, {\dot{q}_2}^2, \\ G_{11} &= m_2 l_2 g \cos(q_1 + q_2) + \big((m_1 + m_2) l_1 g\big)\cos(q_1), \\ G_{22} &= m_2 l_2 g \cos(q_1 + q_2), \\ F_{11} &= f_{v1}\dot{q}_1 + f_{c1} sgn(\dot{q}_1), F_{21} = f_{v2}\dot{q}_2 + f_{c2} sgn(\dot{q}_2), \end{aligned}$$

where $q_1$ and $q_2$are the angles of the robot manipulator joints 1 and 2. The parameter values are $m_1 = 0.8[kg], m_2 = 0.4[kg], l_1 = 0.1[m], l_2 = 0.5[m], g = 9.8\left[\frac{m}{s^2}\right]$.

The friction coefficients are $f_{v1} = 10[Nms], f_{c1} = 10[Nm], f_{v2} = 10[Nms]\ and\ f_{c2} = 10[Nm]$. For the simulation, we set desired angular position $q_d = [q_{1d}, q_{2d}]^T$, $q_{1d} = \frac{\pi}{6}(1 - \cos(1.5\pi t) + \sin(\pi t))$and $q_{2d} = \frac{\pi}{6}(1 - \cos(2\pi t) + \sin(1.5\pi t))$. The adjustable gains are set to $\Lambda = \text{diag}[10,10]$, $\alpha = 10$, $\beta = 200$ and $\epsilon = 0.5$. And the disturbance is given by

$$\begin{cases} [0.1\cos(2t), 0.3\sin(2t)]^T, & t < 30 \\ [0.2\cos(2t), 0.4\sin(3t)]^T, & t < 70 \\ [0.4\cos(4t), 0.3\sin(2t)]^T, & t \geq 70. \end{cases} \quad (10)$$

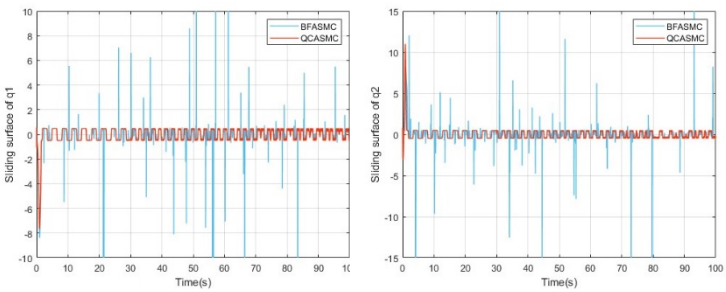



Fig. 1 Siding surface of BFASMC and QCASMC

The simulation result is shown by comparing the sliding surface of BFASMC and QCASMC in Fig. 1. The QCASMC keeps the sliding surface to arbitrarily small vicinity of the zero against the time-varying disturbances. The results of the simulation show that the QCASMC is more robust and stable than the BFASMC.

## V. Conclusions

In this paper, we proposed a new function-based adaptive sliding mode control based on a quasi-convex function. The QCASMC ensures the sliding surface converges to a predefined neighborhood of zero. The QCASMC reduces the overestimation. By comparing QCASMC and BFASMC on the 2-DOF robot manipulator, we show that the QCASMC is more robust and more stable than BFASMC.

## Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ICAN(ICT Challenge and Advanced Network of HRD) program(IITP-2023-2020-0-01832) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation)

## References

[1] V.I. Utkin, Sliding Modes in Control and Optimization. New York: Springer-Verlag, 1992

[2] Huang, Y.J, Kuo, T.C, and Chang, S.H, "Adaptive sliding mode control for nonlinear systems with uncertain parameter" IEEE Transactions on System, Man, and Cybernetics – Part B: Cybernetics, vol. 38, no.2, pp. 534-539, April. 2008.

[3] F. Plestan, Y. Shtessel, V. Bre´geault and A. Poznyak, "New methodologies for adaptive sliding mode control" International Journal of Control, vol. 83, no. 9, pp. 1907-1919, Sep. 2010.

[4] Hussein Obeid, Leonid M. Fridman, Salah Laghrouche, and Mohamed Harmouche, "Barrier function-based adaptive sliding mode control" Automatica, vol. 93, pp. 540-544, July. 2018.

[5] S. Dempe, N. Gadhi, and K. Hamdaoui, "Minimizing the difference of two quasiconvex functions" Optimization Letters, vol. 14, pp. 1765-1779, 2020.

[6] JunMin Park, Wookyong Kwon and PooGyeon Park,"'An Improved Adaptive Sliding Mode Control Based on Time-Delay Control for Robot Manipulators" IEE Transactions on Industrial Electronics, vol. 70, no. 10, pp. 10363-10373, November 2022.

# Two Slope Function-Based Adaptive Sliding Mode Control for IoT that Assist Humans

Dong Hee Seo[1] and Seok Young Lee[1]
[1]*Department of ICT Convergence Engineering, Soonchunhyang University, Asan, Republic of Korea*
*Contact: dable6825@sch.ac.kr

*Abstract*— **For robot manipulators utilized in various facilities, it is important that they execute the desired motion robust to disturbances. This paper proposes an adaptive sliding mode control with a combined adaptive gain that is more robust to time-varying disturbances. The stability of the controller against disturbances is proved by the Lyapunov stability theorem. The simulation results demonstrate that it is more effective than other ASMC.**

## I. INTRODUCTION

Robot manipulators have been utilized in factories [1], medical facilities [2], and other locations that require precision. Especially, since in medical facilities the precision of instrument movement is often directly related to the results of the procedure, robust control algorithms are required to minimize errors and precisely execute the desired motion [2], [3]. However, motion control of robot manipulators is challenging due to nonlinearities, time-varying parameters, unknown disturbances, and modeling uncertainties. These unstable factors harm the control performance of the manipulator and may cause unstableness [4]. To mitigate the problems, a robust control algorithm, called the sliding mode control (SMC) is proposed. SMC is robust to unknown dynamics, and robust stabilization is ideally achieved when the switching gain is greater than an uncertain upper bound. However, in practice, the upper bound is unknown [4]. Thus, the switching gain is selected to be sufficiently large than the uncertainty. This may cause chattering, which can be eliminated by setting the appropriate switching gain [5]. This creates the contradiction of knowing the uncertainty of the unknown. Thus, SMC challenging in real environment applications. To address these problems, adaptive sliding mode control (ASMC) [6], [7] was proposed, that combined adaptive gain. ASMC can adjust the adaptive switch gain regardless of the upper bound of the disturbance. Adaptive gain shows the different effects depending on the designed function. In fact, as stated in [8], the high gain of a monotonically increasing adaptive law can lead to overestimation and cause chattering [9]. To reduce the overestimation, a barrier function-based adaptive sliding mode control (BFASMC) is proposed, which ensures that the sliding surface reaches zero in a finite time [10]. Then, $K_\infty$ function-based adaptive sliding mode control ($K_\infty$ASMC) [11] is proposed, which can ensure the stability of the system even if the sliding variable rapidly changes through the application of BFASMC. Even if the sliding variable changes rapidly, the stability of the $K_\infty$ASMC is ensured if the adaptive control gain is greater than the upper bound of the disturbance. However, the proposed $K_\infty$ASMC function may be susceptible to high disturbance.

## II. PRELIMINARIES

### A. Manipulator system

The dynamics of a typical n-DOF (degree of freedom) manipulator be designed as follow:

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) = u(t) + d \quad (1)$$

where $q \in \mathbb{R}^n$ is a position, $\dot{q} \in \mathbb{R}^n$ is a velocity, $\ddot{q} \in \mathbb{R}^n$ is an acceleration, respectively. $M(q) \in \mathbb{R}^{n\times n}$ is an inertia matrix, $C(q,q) \dot{\in} \mathbb{R}^n$ is Coriolis matrix, $G(q) \in \mathbb{R}^n$ is a gravity force vector, respectively. $u(t) \in \mathbb{R}^n$ is input torque vector. $d \in \mathbb{R}^n$ is unknown disturbance torque from an manipulator. We can summarize the dynamics (1) with respect to $\ddot{q}$ as follow:

$$\ddot{q} = M^{-1}(u(t) - C(q,\dot{q})\dot{q} - G(q) + d) \quad (2)$$

Next, define a system error as follow:

$$e = q_d - q, \quad (3)$$

where $q_d$ represented the desired position trajectory, based on the error dynamics (3), we can define $\dot{e}, \ddot{e}$ as the velocity error, and acceleration error, respectively

### B. Adaptive sliding mode control

In this section, Consider the following first-order system with error $e = q_d - q$,

$$s = \Lambda e + \dot{e}, \quad (4)$$

where $s$ is the sliding variable, $\Lambda \in \mathbb{R}^{n\times n}$ is a positive diagonal matrix. Utilizing the system dynamics (2), the derivative of sliding variable $s$ can be found as follow:

$$\begin{aligned}\dot{s} &= \Lambda e \dot{+} \ddot{e} = \Lambda\dot{e} + \ddot{q}_d - \ddot{q} \\ &= \Lambda\dot{e} + \ddot{q}_d - M^{-1}(u(t) - C(q,\dot{q}) - G(q) + d)\end{aligned} \quad (5)$$

The paper [6] design $u$ the following inputs

$$u = u_0 + u_1, \quad (6)$$

$$u_0 = C + G + M(\ddot{q}_d + \Lambda\dot{e}),$$

$$u_1 = M\hat{K}sgn(s),$$

$$\dot{s} = -M\hat{K}sgn(s) + d, \quad (7)$$

Through the operations in (4) and (5), we can derive the first-order system (6). $\hat{K} \in \mathbb{R}^{n\times n}$ is a gain matrix. $K^* \in \mathbb{R}^{n\times n}$ is a gain at the point where it stops increasing. $\alpha, \beta, \gamma \in \mathbb{R}$ are a

positive adaptive gains. The adaptive law is defined as the formula below

$$\hat{K}(i,i) = \alpha\left(1 - e^{-\beta|s_i|^{\gamma}}\right) + |s_i| \tag{8}$$

$\hat{K}$ increase until the sliding surface reaches zero.

## III. Main Resurlt

### A. Controller Design

In the equation (7), $\hat{K}$ and system with controller are considered:

$$\hat{K} = \alpha\left(1 - e^{-\beta|s|^{\gamma}}\right) + |s|$$
$$u = -\hat{K}sgn(s) \tag{9}$$

### B. Stability analysis

In this subsection, prove the stability of the proposed controller using Lyapunov stability theorem.

$$\hat{K} = \alpha\left(1 - e^{-\beta|s|^{\gamma}}\right) + |s|$$

The stability of the proposed controller is proved by the Lyapunov stability theorem. The Lyapunov function stability equation is as follow:

$$V = \frac{1}{2}s^2$$
$$\dot{V} = s\dot{s}$$
$$= s\left(-\hat{K}sgn(s) + d\right)$$
$$= \left(-\hat{K}|s| + ds\right)$$
$$\leq \left(-\hat{K} + \bar{d}\right)|s| < 0\,,$$

such that

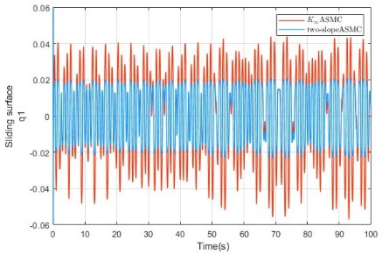
$$\hat{K} > \bar{d} \tag{11}$$

The Lyapunov stability is proved based on [11].

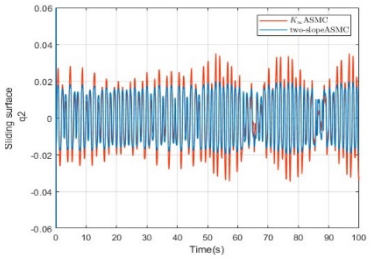
## IV. Simulation Result

The proposed controller is simulated in 2-DOF manipulator dynamics. For the simulation, we set the desired position trajectory $q_d = [q_{1d}, q_{2d}]^T$. The parameter values of $q_d$ are $q_{1d} = \frac{\pi}{6}(1 - \cos(1.5\pi t) + \sin(\pi t))$, and $q_{2d} = \frac{\pi}{6}(1 - \cos(2\pi t) + \sin(1.5\pi t))$, $\Lambda = \text{diag}\,[10, 10]$, Adgustable gain $\beta = 0.98$, $l_1 = 0.1m,,\ l_2 = 0.5m\,, m_1 = 0.8kg, m_2 = 0.4kg, g = 9.8m/s$

$$d(t) = \begin{cases} [0.1\sin(3t)\,, 0.1\sin(3t)]^T & if\ t < 50s \\ [0.2\cos(5t)\,, 0.2\cos(5t)]^T & otherwise \end{cases} \tag{12}$$




(a)




(b)

Fig.1 – Comparison of the sliding variable trajectories by proposed TSASMC and $K_{\infty}$ASMC for each joint. (a) joint1, (b) joint 2.

Fig1 shows the sliding variable trajectories of the two ASMCs. That is, Fig1 clearly shows that the TSASMC has better control performance and is more robust than $K_{\infty}$ function-based ASMC in keeping the sliding surface closer to zero against time-varying disturbance.

## V. Conclusions

In this paper, we proposed an adaptive sliding mode control with an adaptive gain as a combined function. The proposed controller can stabilize the system even in the presence of an unknown disturbance, and it has been verified that it is more stable than the different methods.

## Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ICAN(ICT Challenge and Advanced Network of HRD) program(IITP-2023-2020-0-01832) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation)

## References

[1] CHEBAB, Z. E., et al. Autonomous collaborative mobile manipulators: State of the art. In: *Symposium on Theory of Machines and Mechanisms/UMTS2015/TrISToMM*. 2015.

[2] MITCHELL, Ben, et al. Development and application of a new steady-hand manipulator for retinal surgery. In: *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007. p. 623-629 .

[3] WEINRIB, Harry P.; COOK, John Querin. Rotational technique and microsurgery. *Microsurgery*, 1984, 5.4: 207-212.

[4] BAEK, Jaemin; JIN, Maolin; HAN, Soohee. A new adaptive sliding-mode control scheme for application to robot manipulators. *IEEE Transactions on industrial electronics*, 2016, 63.6: 3628-3637.

[5] LEE, Junyoung; CHANG, Pyung Hun; JIN, Maolin. Adaptive integral sliding mode control with time-delay estimation for robot manipulators. *IEEE Transactions on Industrial Electronics*, 2017, 64.8: 6796-6804.

[6] HUANG, Ying-Jeh; KUO, Tzu-Chun; CHANG, Shin-Hung. Adaptive sliding-mode control for nonlinearsystems with uncertain parameters. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2008, 38.2: 534-539.

[7] PLESTAN, Franck, et al. Sliding mode control with gain adaptation—Application to an electropneumatic actuator. *Control Engineering Practice*, 2013, 21.5: 679-688.

[8] ROY, Spandan, et al. Overcoming the underestimation and overestimation problems in adaptive sliding mode control. *IEEE/ASME Transactions on Mechatronics*, 2019, 24.5: 2031-2039.

[9] BANDYOPADHYAY, Bijnan, et al. Advances in sliding mode control. *Lecture Notes in Control and Information Sciences*, 2013, 440.

[10] OBEID, Hussein, et al. Barrier function-based adaptive sliding mode control. *Automatica*, 2018, 93: 540-544.

[11] SONG, Jiawei; ZUO, Zongyu; BASIN, Michael. New Class $\mathcal {K} _\infty $ Function-Based Adaptive Sliding Mode Control. *IEEE Transactions on Automatic Control*, 2023.

# Effects of Virtual Shopping Budget-Management Training on Executive Function in Healthy Young Adults: A Pilot Study

Ji Yea Kim [1], Si-An Lee[1], Jin-Hyuck Park[2]
[1]*Department of ICT convergence, The Graduate School, Soonchunhyang University, Asan, Republic of Korea*
[2]*Department of Occupational Therapy, College of Medical Science, Soonchunhyang University, Asan, Republic of Korea*
*Contact: jhpark1217@sch.ac.kr, phone +82-41-530-4773

*Abstract*— **Recently, researchers have been exploring various interventions to enhance executive function, a cognitive process that shapes new behaviors and optimizes strategies in unfamiliar situations. One such intervention is virtual reality-based shopping training. However, this method's simplicity and limited effectiveness in engaging executive function are recognized. This study aims to investigate the impact of virtual reality-based shopping budget management training, incorporating budget management and scoring variables, impacts executive function. Eight healthy adult participants took part in this training, engaging in two sessions per week over eight weeks, totalling 16 sessions. This training approach introduced greater complexity in budget management by adding variables to the existing training method and reducing the potential learning effect through four levels of difficulty and varying budget scenarios. The study revealed a significant improvement in executive function, as indicated by the Trail Making Test Part B (TMT-B) and Stroop test, both administered before and after training. Further research is suggested, involving diverse research groups and larger sample sizes.**

## I. INTRODUCTION

"Executive function" refers to a high-level cognitive process composed of subcomponents such as inhibition control, working memory, cognitive flexibility, reasoning, problem-solving, and planning [1, 2]. It is required when deliberately planning and efficiently executing actions [3]. Deficits in executive function can result in delayed initiation or an inability to inhibit automatic responses, especially in situations that require self-regulation and autonomy, such as Instrumental Activities of Daily Living (IADL) [4]. Additionally, it can lead to reduced problem-solving abilities and inappropriate coping strategies [5], resulting in various challenges in everyday life [6].

The goal of rehabilitation, particularly within occupational therapy, is to enhance participation in everyday real-world activities [7]. However, traditional cognitive training is time-consuming and poses safety concerns making it challenging to implement in actual environments [8, 9]. Consequently, ecological validity is lacking [10]. To address these drawbacks, various types of training programs based on virtual reality are being utilized [6, 9, 11-15]. These programs offer the advantage of adjusting situations, frequency, and difficulty levels while allowing training in a safe environment [14]. Furthermore, according to the research by Son & Park (2022), virtual reality-based cognitive training has demonstrated ecological validity.

Recently, training programs, particularly focused on shopping, have been developed to assess and address executive function issues [12, 15, 16]. Shopping is typically considered a representative activity within Instrumental Activities of Daily Living (IADL), and it has been chosen as a meaningful activity for most individuals in their daily lives [17]. Managing and spending money within a given budget while shopping is a complex task that demands intact executive functions [15]. However, existing virtual reality-based shopping training programs consist mainly of simple tasks involving selecting items within a given budget, making it challenging to expect significant improvement in executive function.

Therefore, in this study, we aimed to develop a new program by adding variables to existing shopping training programs, increasing the level of difficulty, and reducing the learning effect. The objective was to investigate the impact of this new program on executive function.

## II. METHODS

### A. Participants

8 healthy young adults were enrolled in this study. The participants had no visual or auditory impairments, were not receiving behavioural interventions for cognitive enhancement, demonstrated proficiency in comprehending verbal instructions, and exhibited no constraints in basic activities of daily living. The exclusion criteria comprised individuals with diagnosed psychiatric or neurological conditions and those having an aversion to computer usage.

### B. Intervention

Virtual shopping budget-management training was conducted over 8 weeks for a total of 16 sessions. Participants received virtual shopping budget-management training through a program created using Scratch (MIT Media Lab, USA). This program is an extension of conventional virtual shopping training programs, featuring added elements of budget management training. The existing program was designed to simulate shopping in a supermarket.

Participants navigated within a virtual supermarket, purchasing items from a shopping list. They had the freedom to navigate the virtual supermarket, selecting and purchasing items by touch. To make a payment, they were required to locate a cash desk, which then prompted the payment screen to appear. Following the payment, a statical screen with feedback on their performance appeared [16]. In another program, participants engaged in the Adapted Four-Item Shopping Task. This task involved four shopping lists, with items positioned in two different aisles. Participants were provided with the lists, eliminating the need to memorize the items [18].

The virtual shopping budget-management program introduced complexity through the incorporation of 'scores' tied to product prices—lower prices corresponded to lower scores, while higher prices correlated with higher scores. Participants were asked to maximize their scores within a predetermined budget while retaining knowledge of their scores. This program places particular emphasis on budget management within the context of executive function training. It offers four levels of difficulty, challenging even young and healthy adults, and utilizes various budget allocations to reduce learning effects (Figure 1,2).


Fig. 1 The virtual shopping budget-management program. Full view of the aisles.


Fig. 2 The virtual shopping budget-management program. A view of some of the shelves.

### C. Measurement

Participants' executive functions were assessed using the Trail Making Test Part B (TMT-B), the Stroop test. For the TMT-B measurement, participants were instructed to draw lines connecting numbers and letters alternately in ascending order [20]. The time it took for participants to complete the TMT-B within 300 seconds was analysed.

The Stroop test typically comprises 2-4 performance tasks. In this study, participants engaged in the "interference color naming" task, where they named the color of words written in ink different from the word's meaning [21]. The Stroop test results were analysed if completed within a maximum of 300 seconds.

### D. Statistical analysis

We analysed all data using the SPSS 22.0 version (SPSS Inc., USA). To compare the within-group pre-post assessment results, the Wilcoxon signed-rank test was employed.

## III. RESULTS

A total of 8 participants (mean age = 20.38 years [SD = 0.92], n = 8 females) were recruited in this study. There were significant differences for both the TMT-B ($p < 0.05$) and Stroop test ($p < 0.05$) (Table 1).

TABLE I
COMPARISON OF EXECUTIVE FUNCTION

| VARIABLES | PARTICIPANTS (n = 8) | Z | *p* |
|---|---|---|---|
| TMT-B (sec) | | | |
| Pre-intervention | 19.01 ± 5.26 | -2.380 | .017* |
| Post-intervention | 15.64 ± 4.88 | | |
| STROOP TEST (sec) | | | |
| Pre-intervention | 87.86 ± 18.94 | -2.380 | .017* |
| Post-intervention | 77.62 ± 12.48 | | |

*p < 0.05, TMT-B: Trail Making Test Part B

## IV. DISCUSSION

This study was designed to investigate the effects of virtual reality shopping budget-management training on executive function. To accomplish this, the participants engaged in 16 sessions, and significant differences in executive function were observed after the training. This result is consistent with previous research findings, demonstrating the effectiveness of virtual shopping training in improving executive function [13, 15].

One significant difference between the existing shopping training and the current study is the emphasis on budget management within the shopping tasks. Existing shopping training was relatively straightforward, primarily involving the task of selecting items within a budget constraint. This simplicity allowed for the implementation of a strategy focused on purchasing the cheapest items, which might not have effectively enhanced executive function. In contrast, the virtual shopping budget-management training introduced variability in prices across different brands, with higher prices yielding more points. To achieve higher scores within the given budget, participants had to engage in more complex mental calculations of adding and subtracting items. Consequently, this training presented a more intricate task compared to previous ones[12, 15, 16], requiring efforts to choose more suitable items within the budget. As the task complexity increased, this training closely resembled the actual shopping environment, encouraging participants to exhibit more rational consumption behavior.

In this study, we observed improved executive function in healthy young adults after completing virtual shopping budget-management training in healthy young adults. This suggests that virtual shopping budget-management training could be effective in enhancing executive function not only in clinical groups such as MCI or stroke patients but also in healthy individuals By emphasizing budget management and offering four stages of varying difficulty, we can provide effective training for diverse participants, including healthy adults.

Additionally, despite the small sample size and nonsignificant results in primary analyses, our research can contribute to the development of effective virtual reality education programs for shopping budget management in the future. Hence, our study can serve as a foundational step towards creating practical virtual reality education for budget management during shopping. Furthermore, these findings can also benefit rehabilitation and elderly care programs. Training in a virtual environment enhances cognitive abilities and daily functioning, with potential for broader societal impact [19].

This study has several limitations. Firstly, the training was conducted on healthy young adults, despite the aim of enhancing executive function. Nevertheless, the observed improvement in executive function in this population suggests the potential applicability of virtual shopping budget-management training to various groups. Secondly, the limited sample size in this study hinders the generalization of the research findings. Thirdly, there was a difference in gender distribution between groups. According to previous studies, gender differences in executive function are not a significant factor, but there might be potential impacts in some aspects of executive function where gender differences are observed [20].

Future research should consider recruiting a larger sample, particularly individuals with executive function impairment.

## V. CONCLUSION

The findings of this study demonstrate the effectiveness of virtual shopping budget-management training in enhancing executive function in healthy young adults. Through VR, it enhances the ecological validity of shopping training and, by focusing on budget management training, increases the difficulty level to closely mimic the real-world environment. Thus, it provides effective training for healthy young adults. These results suggest that virtual shopping budget-management training can serve as an alternative for enhancing executive function in various groups of participants.

## ACKNOWLEDGMENT

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ICAN (ICT Challenge and Advanced Network of HRD) program (IITP-2023-2020-0-01832) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation).

## IRB APPROVAL

The study received approval from the University of Soonchunhyang Institutional Review Board, with the IRB Approval Number: 202302-SB-013.

## REFERENCES

[1] A. Diamond, "Executive functions," *Annu Rev Psychol,* vol. 64, pp. 135-68, 2013, doi: 10.1146/annurev-psych-113011-143750.

[2] S. J. Gilbert and P. W. Burgess, "Executive function," *Current Biology,* vol. 18, no. 3, pp. R110-R114, 2008.

[3] W. A. Gordon, J. Cantor, T. Ashman, and M. Brown, "Treatment of post-TBI executive dysfunction: application of theory to clinical practice," *The Journal of Head Trauma Rehabilitation,* vol. 21, no. 2, pp. 156-167, 2006.

[4] N. Katz and A. Maeir, "Higher-level cognitive functions enabling participation: Awareness and executive functions," 2011.

[5] O. Godefroy, "Frontal syndrome and disorders of executive functions," *J Neurol,* vol. 250, no. 1, pp. 1-6, Jan 2003, doi: 10.1007/s00415-003-0918-2.

[6] J. C. Millan-Calenti *et al.*, "Prevalence of functional disability in activities of daily living (ADL), instrumental activities of daily living (IADL) and associated factors, as predictors of morbidity and mortality," *Arch Gerontol Geriatr,* vol. 50, no. 3, pp. 306-10, May-Jun 2010, doi: 10.1016/j.archger.2009.04.017.

[7] M. Jacoby, S. Averbuch, Y. Sacher, N. Katz, P. L. Weiss, and R. Kizony, "Effectiveness of executive functions training within a virtual supermarket for adults with traumatic brain injury: a pilot study," *IEEE Trans Neural Syst Rehabil Eng,* vol. 21, no. 2, pp. 182-90, Mar 2013, doi: 10.1109/TNSRE.2012.2235184.

[8] P. E. Bailey, J. D. Henry, P. G. Rendell, L. H. Phillips, and M. Kliegel, "Dismantling the "age–prospective memory paradox": The classic laboratory paradigm simulated in a naturalistic setting," *Quarterly Journal of Experimental Psychology,* vol. 63, no. 4, pp. 646-652, 2010.

[9] R. Kizony, N. Josman, N. Katz, D. Rand, and P. Weiss, "Virtual reality and the rehabilitation of executive functions: An annotated bibliography," *Isr. J of Occup. Ther,* vol. 17, no. 2, pp. E47-61, 2008.

[10] N. Josman and N. Katz, "Relationships of categorization on tests and daily tasks in patients with schizophrenia, post-stroke patients and healthy controls," *Psychiatry Res,* vol. 141, no. 1, pp. 15-28, Jan 30, 2006, doi: 10.1016/j.psychres.2004.03.015.

[11] C. Son and J. H. Park, "Ecological Effects of VR-Based Cognitive Training on ADL and IADL in MCI and AD patients: A Systematic Review and Meta-Analysis," *Int J Environ Res Public Health,* vol. 19, no. 23, Nov 29, 2022, doi: 10.3390/ijerph192315875.

[12] J. H. Park, "Does the virtual shopping training improve executive function and instrumental activities of daily living of patients with mild cognitive impairment?," *Asian J Psychiatr,* vol. 69, p. 102977, Mar 2022, doi: 10.1016/j.ajp.2021.102977.

[13] J. H. Park, "Effects of Cognitive-Physical Dual-Task Training on Executive Function and Activity in the Prefrontal Cortex of Older Adults with Mild Cognitive Impairment," *Brain Neurorehabil,* vol. 14, no. 3, p. e23, Nov 2021, doi: 10.12786/bn.2021.14.e23.

[14] P. Mesa-Gresa, H. Gil-Gomez, J. A. Lozano-Quilis, and J. A. Gil-Gomez, "Effectiveness of Virtual Reality for Children and Adolescents with Autism Spectrum Disorder: An Evidence-Based Systematic Review," *Sensors (Basel),* vol. 18, no. 8, Aug 1 2018, doi: 10.3390/s18082486.

[15] S. Y. Nir-Hadad, P. L. Weiss, A. Waizman, N. Schwartz, and R. Kizony, "A virtual shopping task for the assessment of executive functions: Validity for people with stroke," *Neuropsychol Rehabil,* vol. 27, no. 5, p. 808-833, Jul 2017, doi: 10.1080/09602011.2015.1109523.

[16] S. Zygouris *et al.*, "Can a virtual reality cognitive training application fulfill a dual role? Using the virtual supermarket cognitive training application as a screening tool for mild cognitive impairment," *J Alzheimers Dis,* vol. 44, no. 4, pp. 1333-47, 2015, doi: 10.3233/JAD-141260.

[17] J. L. Thompson, G. Bentley, M. Davis, J. Coulson, A. Stathi, and K. R. Fox, "Food shopping habits, physical activity and health-related indicators among adults aged⩾ 70 years," *Public health nutrition,* vol. 14, no. 9, pp. 1640-1649, 2011.

[18] D. Rand, N. Katz, and P. L. Weiss, "Evaluation of virtual shopping in the VMall: Comparison of post-stroke participants to healthy control groups," Disability and rehabilitation, vol. 29, no. 22, pp. 1710-1719, 2007.

[19] A. C. M. Bauer and G. Andringa, "The potential of immersive virtual reality for cognitive training in elderly," Gerontology, vol. 66, no. 6, pp. 614-623, 2020.

[20] N. M. Grissom and T. M. Reyes, "Let's call the whole thing off: evaluating gender and sex differences in executive function," (in eng), Neuropsychopharmacology, vol. 44, no. 1, pp. 86-96, Jan 2019, doi: 10.1038/s41386-018-0179-5.

# Performance of ChatGPT on the National Korean Occupational Therapy Licensing Examination

Jin-Hyuck Park[1*]

[1]*Department of Occupational Therapy, College of Medical Science, Soonchunhyang University, Asan, Republic of Korea*
*Contact: jhpark1217@sch.ac.kr, phone +82-41-530-4773

*Abstract*—**ChatGPT is an artificial intelligence-based large language model (LLM). ChatGPT has been widely applied in medicine, but its application in occupational therapy has been lacking. This study examined the accuracy of ChatGPT on the National Korean Occupational Therapy Licensing Examination (NKOTLE) and investigated its potential for application in the field of occupational therapy. ChatGPT 3.5 was used on the five years of the NKOTLE with Korean prompts. Multiple choice questions were entered manually by three dependent encoders, and scored according to the number of correct answers. During the most recent five years, ChatGPT did not achieve a passing score of 60% accuracy and exhibited interrater agreement of 0.6 or higher. ChatGPT could not pass the NKOTLE but demonstrated a high level of agreement between raters. Even though the potential of ChatGPT to pass the NKOTLE is currently inadequate, it performed very close to the passing level even with only Korean prompts.**
.

## I. INTRODUCTION

In recent decades, deep learning advancements have revolutionized artificial intelligence (AI) across various industries.1-3 Notably, AI's ability to accurately classify traditional audio, images, and text data has enabled object categorization for photos and human-level text translation.1-3 Recently, there has been significant interest in large language models (LLMs)-based AIs with unique capability to generate responses based on natural language input.

Unlike AIs limited to domain-specific data in a certain field, LLMs-based AIs can analyze non-domain-specific data. This characteristic eliminates the necessity of creating highly domain and problem-specific training data, thus enhancing their performances.4 Given these advantages, the medical field has begun exploring the application of LLMs-based AIs for personalized healthcare, including diagnosis, clinical image analysis, and disease prediction.5

Consequently, there has been a surge of interest in leveraging Chatbot Generative Pre-trained Transformer (ChatGPT), a prominent LLM-based AI, within the medical domain. ChatGPT is fine-tuned by LLMs that are trained based on text data from the Internet via reinforcement and supervised learning ways.6 Several studies have examined ChatGPT's performance in medical licensing examinations to assess its ability to interact with patients based on medical knowledge. In a previous study, ChatGPT was employed in the United States Medical Licensing Examination (USMLE) and achieved a passing-level accuracy (60%) for medical-related natural language questions.7 The clinical significance of ChatGPT lies in its ability to achieve a passing-level performance on the USMLE without needing for a professional human trainer. Furthermore, other studies demonstrated that ChatGPT can also pass the Japanese Medical Licensing Examination (JMLE) and Taiwan pharmacist licensing examination.8,9 In other words, ChatGPT could serve as a tool for medical assistance or self-study for medical students across different countries,10 even in non-English-language studies where prompts were presented in both native languages and English translations.8,9

Although there have been some studies on ChatGPT for pharmacist and nurse licensing examinations,9,11 the performance of ChatGPT has been primarily evaluated in the context of medical licensing examinations. Therefore, its applicability in other medical fields remains uncertain. Physicians and occupational therapists (OTs) both play crucial roles in the rehabilitation system. However, their specialties and responsibilities differ. Physicians primarily focus on diagnosing and treating illness, while OTs are heavily involved in facilitating clients' engagement in occupations. Specifically, OTs address various aspects of clients' performances such as physical, cognitive, psychological, and sensory-perceptual factors to support their engagement in occupations.12

Given the diversity of medical fields, this study aimed to assess the performance of ChatGPT using questions from the National Korean Occupational Therapy Licensing Examination (NKOTLE). The NKOTLE encompasses all knowledge domains essential for occupational therapists.13 The difficulty and complexity of the NKOTLE questions are highly standardized and regulated by a panel of experts, making them suitable for AI testing. These questions are well-established, with very stable raw scores and psychometric properties over the past five years.14 In addition, as the NKOTLE questions are exclusively in a multiple-choice, and text-oriented format, they provide a challenging assessment for ChatGPT. However, no previous studies have reported the performance of ChatGPT on the NKOTLE.

Therefore, this study aimed to determine if ChatGPT could successfully pass the last five years of the NKOTLE, offering quantitative feedback on its performance and evaluating its potential for application in the field of occupational therapy,

similar to other medical domains. Additionally, unlike prior works that utilized prompts in both native language and English translations, this study sought to verify the feasibility of using a non-English language-based prompt by presenting all questions exclusively in Korean.

## II. METHODS

### A. Artificial Intelligence

ChatGPT is an advanced language model developed by OpenAI, located in San Francisco, CA, USA. This model utilizes self-attention mechanisms and extensive training data to generate coherent and contextually appropriate responses in natural language within a conversational context. It excels in handling long-range dependencies, ensuring that its generated responses are well-connected and relevant. Unlike other conversational systems or Chatbots that have access to external sources of information, such as internet searches or databases, ChatGPT is a self-contained server-based model. Consequently, all responses it produces are generated within the model itself, based on abstract relationships between words or "tokens" within its neural network.7 In this study, the freely available version 3.5 of ChatGPT was utilized.

### B. Inputs

Public-available test questions from NKOTLE-2018 to NKOTLE-2022 were from the official NKOTLE website. Due to the lack of access to previous data, only questions from the past five years were utilized for this study. The NKOTLE is taken by students who are planning to graduate from a three- or four-year occupational therapy program in South Korea. The NKOTLE consists of three units. The first unit has two sub-units. The first sub-unit assesses basic knowledge of occupational therapy including anatomy, physiology, and public health with 70 questions, and the second sub-unit tests the medical service act with 20 questions. The second unit tests specialized occupational therapy knowledge such as neurological, musculoskeletal, and psychiatric occupational therapy with 100 questions. The third unit assesses clinical reasoning through a 50-question paper-and-pencil practical examination that provides illustrations and hypothetical clinical data. The passing threshold is 60% or more correct answers for each unit, and less than 40% of correct answers for each unit will fail regardless of the overall percentage of correct answers. Students who pass this examination are licensed by the Ministry of Health and Welfare of South Korea and registered as occupational therapists.13

In this study, only the first and second units of the NKOTLE problems were utilized due to copyright restrictions preventing access to the third unit, which is not disclosed on the NKOTLE website. Furthermore, questions of the third unit primarily contain images or graphs that depict medical conditions of virtual cases, making it unsuitable for prompt input in the CHATGPT 3.5 version. Both the first and second units consist entirely of multiple-choice questions with a single answer, where forced justification is not required.

To ensure that input data used in the study were representative, 20 test questions per year were randomly sampled from the first unit and second unit questions. It was verified that these NKOTLE questions were not indexed in Google after January 1, 2018. Subsequently, it was confirmed that these questions from the first and second units did not include images or graphs. After filtering, a total of 950 questions (190 questions per year) were advanced to encoding.

### C. Encoding

This study encoded by reproducing the original NKOTLE questions verbatim. To ensure a single correct answer, the following prompt was consistently encoded with each question: “Which of the following best represents the most appropriate answer?” Three encoders independently encoded the prompt. A new chat session was started in ChatGPT for each entry to reduce memory retention bias. If ChatGPT failed to provide an answer for a question in terms of the answer choice number or text, the question was re-encoded up to three times.

### D. Accuracy and interrater agreement

The three encoders independently evaluated the accuracy of ChatGPT based on the criterion for determining the correct answer. According to this criterion, ChatGPT should either present the correct answer number or the text corresponding to the correct number as provided in the NKTOLE. The correct answer for each question was coded in an Excel file and shared with an independent examiner. The independent examiner analyzed the percentage of correct answers determined by encoders, compared it against the passing criteria, and assessed the inter-rater agreement between encoders. A schematic of the study flow is provided in Figure 1.

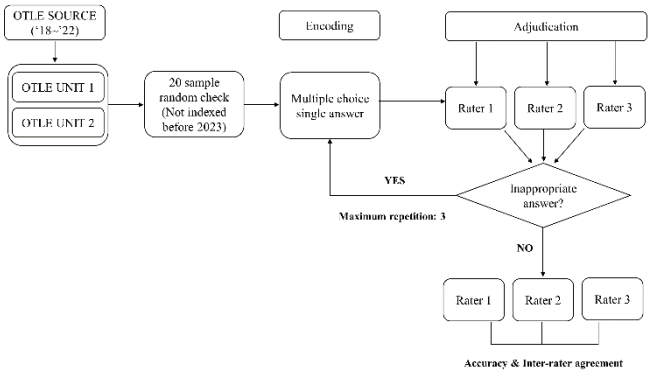



**Fig. 1.** Flowchart for sourcing, encoding, and adjudicating results.

### E. Statistical analyses

All data were analyzed using IBM SPSS Statistics version 22.0. The Fleiss’ kappa statistics were used to compute inter-rater agreement on the accuracy of ChatGPT on the NKOTLE. The Spearman correlation test was conducted to confirm the correlation between accuracy and inter-rater agreement. Statistical significance was set at $p < 0.05$.

Table I
ChatGPT accuracy across the past five years of the NKOTLE

| NKOTLE | | Accuracy (%) First rater | Second rater | Third rater | Avg | Benchmark (>60.0) | Fleiss kappa |
|---|---|---|---|---|---|---|---|
| 2018 | Unit 1-1 | 67.1 | 70.0 | 67.1 | 68.1 | Fail (59.2) | .744*** |
| | Unit 1-2 | 25.0 | 20.0 | 25.0 | 23.3 | | .756*** |
| | Unit 2 | 73.0 | 63.0 | 60.0 | 65.3 | | .719*** |
| 2019 | Unit 1-1 | 68.6 | 64.3 | 65.7 | 66.2 | Fail (53.2) | .833*** |
| | Unit 1-2 | 40.0 | 35.0 | 35.0 | 36.7 | | .656** |
| | Unit 2 | 51.0 | 58.0 | 61.0 | 56.7 | | .741*** |
| 2020 | Unit 1-1 | 78.6 | 65.7 | 67.1 | 70.5 | Fail (59.2) | .845*** |
| | Unit 1-2 | 45.0 | 50.0 | 50.0 | 48.3 | | .831*** |
| | Unit 2 | 61.0 | 57.0 | 58.0 | 58.7 | | .813*** |
| 2021 | Unit 1-1 | 65.7 | 60.0 | 57.1 | 61.0 | Fail (59.3) | .672*** |
| | Unit 1-2 | 75.0 | 40.0 | 50.0 | 55.0 | | .602** |
| | Unit 2 | 58.0 | 61.0 | 67.0 | 62.0 | | .751*** |
| 2022 | Unit 1-1 | 54.3 | 71.4 | 70.0 | 65.2 | Fail (58.2) | .807*** |
| | Unit 1-2 | 50.0 | 35.0 | 60.0 | 48.3 | | .660** |
| | Unit 2 | 68.0 | 56.0 | 59.0 | 61.0 | | .813*** |

## III. Results

### A. Accuracy

The performance (correct, incorrect, and undetermined answers) of ChatGPT on the NKOTLE by year is presented in Figure 2. In the 2018 NKOTLE, the average accuracy of the three units was 52.2%. In the 2019 NKOTLE, the average accuracy of the three units was 53.2%. In the 2020 NKOTLE, the average accuracy of the three units was 59.2%. In the 2021 NKOTLE, the average accuracy of the three units was 59.3%. In the 2022 NKOTLE, the average accuracy of the three units was 58.2%. For both 2020 and 2021 NKOTLEs, ChatGPT showed a near-passing accuracy. Undetermined answers were counted as incorrect and then the accuracy of ChatGPT was calculated. Table 1 shows each rater's ChatGPT accuracy across the past five years of the NKOTLE.

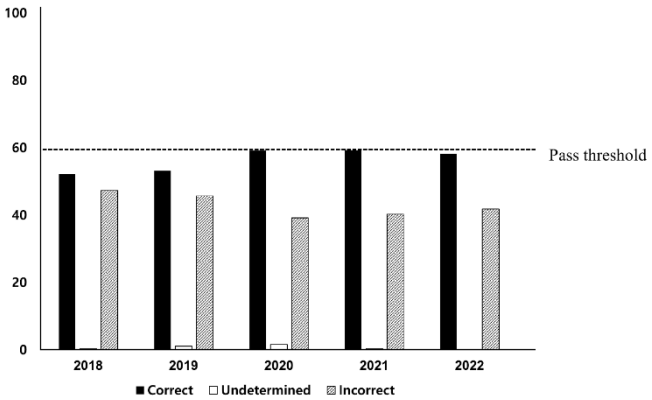



**Fig. 2**. Accuracy of ChatGPT on the NKOTLE.

### B. Interrater agreement

Regarding the inter-rater reliability of ChatGPT between raters, the Fleiss kappa ranged from 0.602 to 0.845 (p's < 0.01)

(Table 1), suggesting that the inter-rater reliability of ChatGPT on the NKOTLE was acceptable.

## IV. Discussion

Our study aimed to examine the feasibility of ChatGPT on the NKOTLE. The current findings demonstrated that ChatGPT could not pass the examination. Nevertheless, ChatGPT was close to the passing score with only a few more correct answers in some years of the NKOTLE. To sum up, the proficiency and ability of ChatGPT to interpret questions pertaining to the NKOTLE are currently inadequate.

ChatGPT emergence marks a significant advancement in natural language processing.15 Its continuous growth promises profound impacts across various sectors such as business, medicine, education, and entertainment.16 By leveraging its extensive language data learning, ChatGPT crafts human-like text and can revolutionize education.15 Particularly in medical studies, it offers personalized learning, aids in exam inquiries, and enhances student engagement.17 In this vein, while ChatGPT has been evaluated for exams in a variety of healthcare fields, to the best of our knowledge, this is the first study to determine the performance of ChatGPT on an occupational therapy licensing examination.

In this study, across the five years of the NKOTLE, ChatGPT's score did not meet the passing requirement. According to statistics, the average pass rate of the NKOTLE across the five years was 88.8%.18 Compared to occupational therapy students who have undergone traditional 3 or 4-year education, ChatGPT's performance is currently insufficient. Considering that ChatGPT 3.5 works well with multiple-choice questions and that NKOTLE questions used in this study are all multiple-choice questions, a high accuracy of ChatGPT was expected. However, ChatGPT did not pass the 2018-2022 examinations. This misalignment with the expectation might be due to the following reasons. Firstly, there are differences in medical laws and policies between South Korea and the States. Specifically, when comparing accuracy by units, it was found that the accuracy of UNIT 1-2 was consistently lower than that of other units. This was because UNIT 1-2 consisted of questions related to medical laws and policies in South Korea, which ChatGPT might lack training on.19 This is in contrast to prior studies showing that ChatGPT came close to a passing score in the English exam.9 Indeed, English is the most resource-rich language in many applications of natural language processing.20 In other words, ChatGPT performs better in English in other languages,21 supporting our assumptions. Consequently, ChatGPT performed relatively well in UNIT 1-1 and UNIT 2 as these units cover subjects such as anatomy, physiology, and specialized occupational therapy areas, which are less than affected by language or national conditions. The accuracy for these units is similar to that of ChatGPT applied to licensing examinations in English or English-speaking cultures.7-9,11 Secondly, all questions in the NKOTLE required selection of the best answer, while some questions had multiple-choice answers provided by ChatGPT, which could be regarded as suboptimal rather than incorrect in clinical practice,19 which is consistent with a previous study.19 Indeed, in a previous study, since ChatGPT sometimes did not understand multiple-choice questions, ChatGPT selected 2 or more answers rather than choosing the best answer.19 These findings suggest that ChatGPT is not yet trained to choose only the best option.

On the other hand, across the five years of the NKOTLE, a high agreement between three raters was observed, indicating that ChatGPT could provide reliable outputs regardless of who would use it. Given that ChatGPT can produce different outputs depending on prompts, this study used uniform prompts. This suggests that using controlled prompts can produce reliable results, supported by a previous study.8,11,22 A previous study showed that inter-rater agreement is positively correlated with accuracy.7 Therefore, it was assumed that the low accuracy was heavily due to missing information rather than over-commitment to incorrect answer choices. In contrast, in this study, there was no significant correlation between accuracy and inter-rater agreement. Thus the low accuracy of ChatGPT on the NKOTLE could be attributed to incorrect answer choices, which is different from the findings of a previous study.7

Justifications for correct answers were not investigated as they did not require justification. However, given the low accuracy of ChatGPT on the NKOTLE, ChatGPT is not sufficiently robust at present to be utilized for the educational demands of occupational therapy students. However, the findings of this study should not be generalized to other subjects or fields as ChatGPT's knowledge will rapidly improve in response to user feedback, leading to different outcomes of subsequent trials using the same questions.8,11,16 Therefore, future studies should investigate whether ChatGPT can also be used for educational purposes to support the process of writing explanations for occupational therapy-related questions, reducing the effort of occupational therapy students.7

While most of the previous studies reported ChatGPT's performance using only English or both native language and English translation as a prompt, this study confirmed similar levels of accuracy to the previous studies,7-9,11 even though only Korean prompts were used. The current findings are especially interesting considering that the training data of LLMs are centered on English,4,23 suggesting that even in non-English-speaking countries, ChatGPT could be utilized in its native language as long as questions are free from language and national conditions. Nevertheless, the impact of English as a prompt could not be ignored as a previous study reported the superiority of English-translated prompts over native-language prompts. This might be overcome to some extent in the future by using extensions such as automatic English translations.24

This study has some limitations. Firstly, we assumed that this study applied ChatGPT to the NKOTLE for the first time. However, this could not be verified. Secondly, due to the lack of disclosure, this study was unable to verify ChatGPT's performance on UNIT 3. Thus, it was not possible to confirm the final KNOTLE pass. Nevertheless, it is currently impossible to verify ChatGPT 3.5 because UNIT 3 contains several images, figures, or tables, which are not interpreted by ChatGPT 3.5. Finally, to improve ChatGPT's performance, this study used a simple prompting method instead of using the original question as a prompt. However, the accuracy might be lower due to the lack of optimized prompt engineering.8 Nonetheless, the results of this study are highly reproducible as this study confirmed ChatGPT performance with an almost exact replication of the

actual question format. However, in the future, it is necessary to prompt ChatGPT so that it does not give incorrect answers when it is unsure of the answer.

## V. CONCLUSION

In conclusion, ChatGPT's knowledge in answering the NKOTLE is not yet comparable to that of occupational therapy students in Korea. However, its ability continues to evolve. We expect higher performance of ChatGPT with prompt engineering and English translation prompts in the future. Therefore, in the near future, professors and students of occupational therapy need to pay attention to the potential of AI chatbots and consider their applications in learning and teaching methods.

## REFERENCES


[1] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* Las Vegas, NV, USA, 27-30 June 2016, pp.2818–2826.

[2] Zhang W, Feng Y, Meng F, et al. Bridging the gap between training and inference for neural machine translation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy, 28 July-2 Aug 2019, pp.4334–4343.

[3] Wang W, and Siau K. Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda. *J Database Manag* 2019; 30: 61–79.

[4] Brown T. Mann B, Ryder N, et al. Language models are few-shot learners. *Adv Neural Inf Process Syst* 2020; 33: 1877–1901.

[5] Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare* 2023; 11: 887.

[6] Castelvecchi D. *Are ChatGPT and AlphaCode going to replace programmers?*. https://www.nature.com/articles/d41586-022-04383-z (2022, accessed 1 Aug 2023).

[7] Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS Digit Health* 2023; 2: e0000198.

[8] Kasai J, Kasai Y, Sakaguchi K, et al. Evaluating GPT-4 and ChatGPT on Japanese medical licensing examinations. *arXiv [csCL]*, https://arxiv.org/abs/2303.18027 (2023).

[9] Wang YM, Shen HW, and Chen TJ. Performance of ChatGPT on the Pharmacist Licensing Examination in Taiwan. *J Chin Med Assoc* 2023; 86: 653–658.

[10] Nisar S and Aslam MS. Is ChatGPT a good tool for T&CM students in studying pharmacology?. https://doi.org/10.2139/ssrn.4324310 (2023, accessed 1 Aug 2023).

[11] Taira K, Itaya T, and Hanada A. Performance of the Large Language Model ChatGPT on the National Nurse Examinations in Japan: Evaluation Study. *JMIR Nursing* 2023; 6: e47305.

[12] Boop C, Cahill SM, Davis C, et al. Occupational therapy practice amework: Domain and process fourth edition. *Am J Occup Ther* 2020; 74: 1–84.

[13] Information on the National Korean Occupational Therapy Licensing Examination. Korea Health Personnel Licensing Examination Institute, https://www.kuksiwon.or.kr/subcnt/c_2015/1/view.do?seq=7&itm_seq=13 (accessed 1 Aug 2023).

[14] Performance data. Korea Health Personnel Licensing Examination Institute, https://www.kuksiwon.or.kr/peryearPass/list.do?seq=13&srchWord=13 (accessed 1 Aug 2023).

[15] Biswas S. ChatGPT and the future of medical writing. *Radiology* 2023; 307: e223312.

[16] Hacker P, Engel A, and Mauer M. Regulating ChatGPT and other large generative AI models. *arXiv [csCL]*, https://arxiv.org/abs/2302.02337 (2023).

[17] Jeblick K, Schachtner B, Dexl J, et al. Chatgpt makes medicine easy to swallow: An exploratory case study on simplified radiology reports. *arXiv [csCL]*, https://arxiv.org/abs/2212.14882 (2022).

[18] Performance data. Korea Health Personnel Licensing Examination Institute, https://www.kuksiwon.or.kr/news/brd/m_54/view.do?seq=453&&itm_seq_1=0&&itm_seq_2=0 (accessed 1 Aug 2023).

[19] Huh S. Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: A descriptive study. *J Educ Eval Health Prof* 2023; 20: 1516081869 (2023).

[20] Névéol A, Dalianis H, Velupillai S, et al. Clinical natural language processing in languages other than English: opportunities and challenges. *J Biomed Semantics* 2018; 9: 1–13.

[21] Hu J, Ruder S, Siddhant A, et al. Xtreme: A massively multilingual multi-task benchmark for evaluating cross-lingual generalization. In *Proceedings of the 39th International Conference on Machine Learning.* Baltimore, Maryland, USA, 17-23 July 2022, pp.4411–4421.

[22] Liu Y, Gu J, Goyal N, et al. Multilingual denoising pre-training for neural machine translation. *Trans Assoc Comput Linguist* 2020; 8: 726–742.

[23] Zhang S, Roller S, Goyal N, et al. Opt: Open pre-trained transformer language models. *arXiv [csCL]*, https://arxiv.org/abs/2205.01068 (2022).

[24] Fang, C. et al. How does ChatGPT4 perform on Non-English National Medical Licensing Examination? An Evaluation in Chinese Language. *medRxiv* https://www.medrxiv.org/content/10.1101/2023.05.03.23289443v1 (2023)

# Aortic Annulus Detection based on Deep Learning for Transcatheter Aortic Valve Replacement using Dual-energy Cardiac Computed Tomography

Yongwon Cho[1,2], Soojung Park[1], Sung Ho Hwang[1], Minseok Ko[1], Jaemin Shim[3], Jing-il Choi[3], Young-Hoon Kim[3], and Yu-Whan Oh[1]

[1]*Department of Radiology, Korea University Anam Hospital, Korea University College of Medicine, 73, Goryeodae-ro, Seongbuk-gu, Seoul 02841, Republic of Korea*
[2]*AI center, Korea University Anam Hospital, Korea University College of Medicine, 73, Goryeodae-ro, Seongbuk-gu, Seoul 02841, Republic of Korea*
[3]Department of Cardiology, Korea University Anam Hospital, Seoul 02841, Republic of Korea

This study aimed to propose a deep learning framework to automatically detect the complex structure of aortic annulus plane on cardiac dual-energy computed tomography (DECT) for transcatheter aortic valve replacement (TAVR) procedure. Annulus Detection Permuted Adain network (ADPANET) based on 3D U-net architecture [1][2][3] (fig. 1). was developed to detect and localize the aortic annulus plane on cardiac DECT. Patients (N = 71) for TAVR between January 2017 and July 2020 from the tertiary medical center were enrolled. Training, tuning and test sets (7:1:2) were used to construct the deep learning model. The root mean square error (RMSE) and Dice similarity coefficient (DSC) were used to evaluate the performance of ADPANET to detect the aortic annulus plane. RMSE and DSC values for aortic annulus plane using ADPANet were 55.078 ± 35.794 and 0.496 ± 0.217, respectively (fig. 2 and 3). Our deep learning framework was feasibly able to detect the 3D complex structure of aortic annulus plane on cardiac DECT for TAVR procedure.

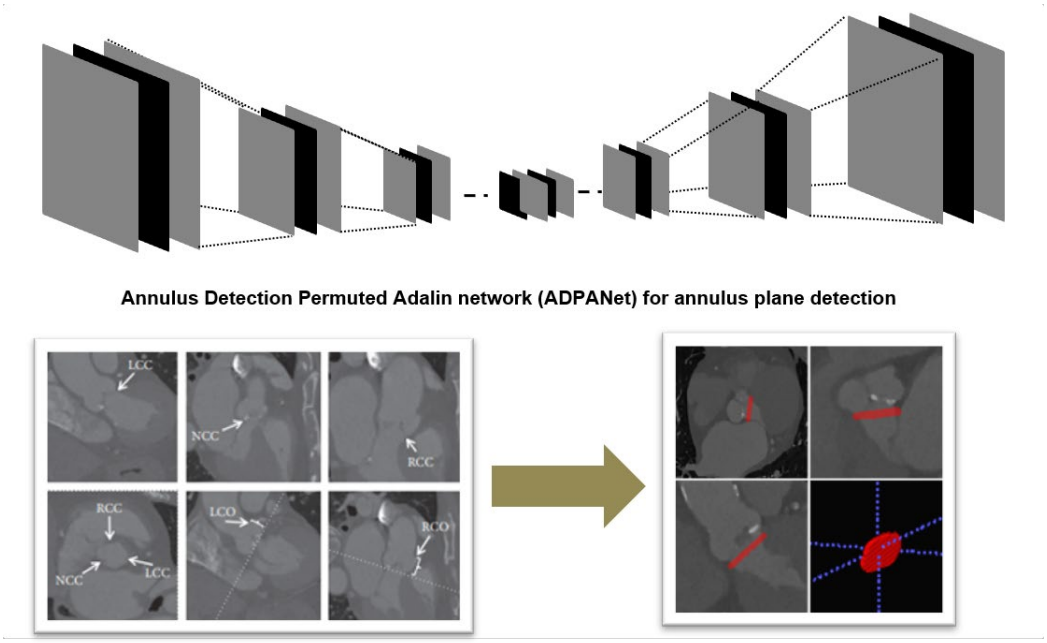



**Fig.1 –** Overview of study

| Measurements | ADPANet | Transformer-based 3D U-net | Original 3D U-net |
|---|---|---|---|
| RMSE | 55.078 ± 35.794 | 58.862 ± 39.538 | 69.656 ± 65.230 |
| DSC | 0.496 ± 0.217 | 0.417 ± 0.2610 | 0.495 ± 0.259 |

**Fig.2 –** Dice similarity coefficient and the root mean square error for detection of the aortic annulus plane using ADPANet, original 3D U-net, and transformer-based 3D U-net for test dataset (11 cases)

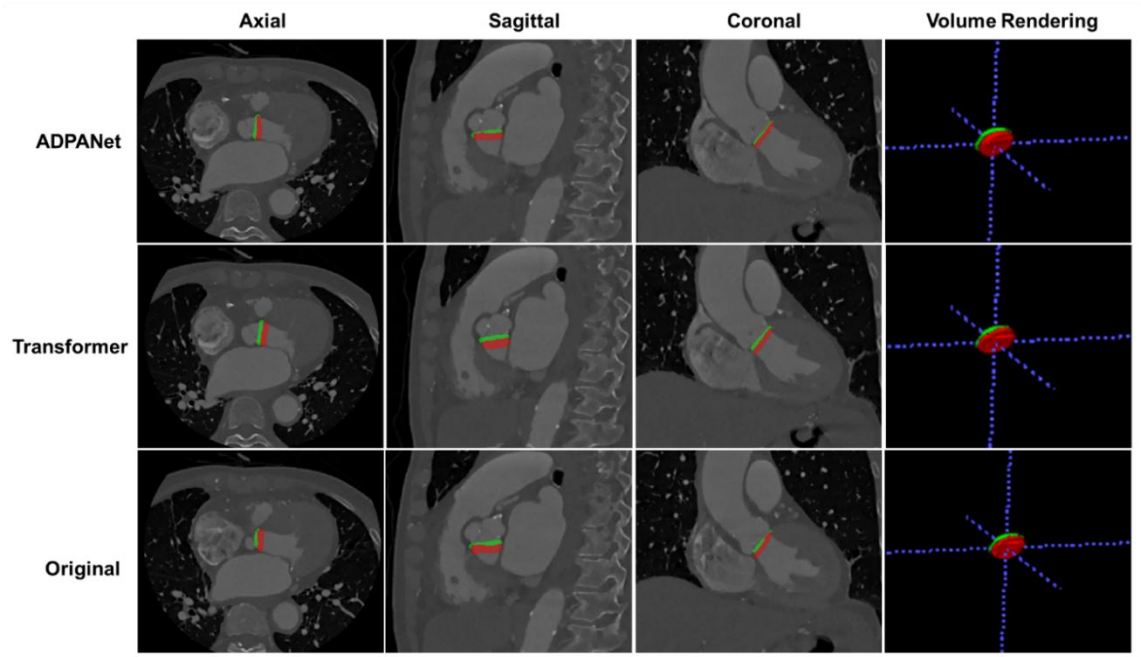



**Fig.3 –** Dice similarity coefficient and the root mean square error for detection of the aortic annulus plane using ADPANet, original 3D U-net, and transformer-based 3D U-net for test dataset (11 cases)

## Funding

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1I1A1A01071600, 2021K1A3A1A88101097) as United Kingdom/South Korea collaboration funding and (2021R1F1A1055272).

## Acknowledgments

We would like to thank the Advanced Medical Imaging Institute in the Department of Radiology, Korea University Anam Hospital in the Republic of Korea, and the researchers for providing software, datasets, and various forms of technical support.

References

[1] Xie Y., Zhang J., Shen C., Xia Y.: CoTr: Efficiently bridging CNN and transformer for 3D medical image segmentation. arXiv preprint arXiv:2103.03024 (2021).

[2] Isensee F., Jaeger P.F., Kohl S.A., Petersen J., Maier-Hein K.H.: nnU-net: Breaking the spell on successful medical image segmentation. Arxiv preprint arXiv:1904.08128 (2019).

[3] Nuriel O., Benaim S., Wolf L.: Permuted AdaIN: reducing the bias towards global statistics in image classification. arXiv preprint arXiv:2010.05785, (2021).

# Effects of Virtual Shopping Budget-Management Training on Brain Activation in Healthy Young Adults: A Pilot Study

Si-An Lee[1], Ji Yea Kim[1], Jin-Hyuck Park[2]
[1]*Department of ICT convergence, The Graduate School, Soonchunhyang University, Asan, Republic of Korea*
[2]*Department of Occupational Therapy, College of Medical Science, Soonchunhyang University, Asan, Republic of Korea*
*Contact: jhpark1217@sch.ac.kr, phone +82-41-530-4773

*Abstract*— **To date, budget-management in virtual shopping training has not been given much importance. The main objective of this study was to investigate the effects of virtual shopping budget-management training on brain activation of dorsolateral prefrontal cortex (dlPFC) area. 16 participants were randomly assigned to the experimental group that received virtual shopping budget-management training or the waitlist control group for a total of 16 sessions. To examine the effects of virtual shopping budget-management training on brain activation, functional near-infrared spectroscopy (fNIRS) was measured during the Trail Making Test Part B(TMT-B) and the Stroop test. After the 16 sessions, there was no significant difference in HbO2 for the TMT-B ($p > 0.05$) and the Stroop test ($p > 0.05$). These results suggest that virtual shopping budget-management training might not be effective in changing the brain structure of healthy young adults.**

## I. INTRODUCTION

Executive functions consist of high-level cognitive processes that facilitate new behavioral patterns and optimize approaches to unfamiliar environments [1]. Impairments in executive function compromise the independence in instrumental activities of daily living (IADL), such as shopping [2]. Therefore, enhancing executive function can be a rehabilitative goal for promoting independent IADL.

Conventional cognitive interventions aimed at enhancing executive functions primarily focus on improving sub-components of executive function through demanding cognitive training with tabletop activities and computerized cognitive training programs, which limits the integration of overall executive functions [3]. Moreover, since their contents are far from real-world, it is difficult for their effects to be transferred to the subject's daily life [4]. Therefore, cognitive training for executive function should consider ecological validity, rather than focusing on sub-elements of executive function [5].

Virtual reality (VR) is being widely applied to ensure ecologically validated cognitive training. It has the advantage of providing cognitive training in a safe environment, personalizing the situations, tasks, and difficulty of training, and providing feedback by tracking and analyzing the participant's functional improvement [6]. Recently, virtual shopping training has been developed to assess and treat executive function impairment [7-9]. This training consists of having the subject remember shopping items, navigate the inside of the supermarket, and find and purchase them on the shopping list within their budget [7]. Spending and managing money within a budget are complex IADLs that heavily depend on intact executive function [9]. However, in traditional VR-based shopping training, budget management typically involves the straightforward selection of products within a predetermined budget, rather than the creation of an optimal purchasing strategy. This approach might not fully engage executive function [10]. Accordingly, virtual shopping budget-management training might yield superior outcomes through the optimization of executive function compared to traditional VR-based shopping training.

On the other hand, previous studies have relied on paper-and-pencil-based executive function assessments to evaluate training's effect [8, 11]. These assessments, however, have limitations in objectively substantiating the basis for training due to their inability to observe clear changes in brain function [8]. Brain imaging techniques such as functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and functional near-infrared Spectroscopy (fNIRS) have been used to measure changes in brain function, which could be evidence of cognitive training efficacy. Of brain imaging techniques, fNIRS non-invasively measures brain activity using changes in light absorption in the brain [12, 13]. fNIRS has the advantages of portability, movement tolerability, and safety of use compared to other neuroimaging modalities [14].

Therefore, the aim of this study is to employ fNIRS to investigate changes in brain activation, thereby enhancing our understanding of the impact of virtual shopping budget-management training on executive functions. Through this investigation, we seek to facilitate improved IADL.

## II. METHODS

### A. Design

This study was a pilot study, and all participants were randomly assigned, with 8 participants in each, to either the experimental group or the wait-list control group using the Python computer language. The intervention consisted of 16 training sessions conducted twice a week over 8 weeks. This study was approved by the Institutional Review Board of the University of Soonchunhyang (protocol code 202302-SB-013).

### B. Participants

16 healthy young adults participated in this study. Participants included in this study were individuals who had no visual or auditory impairments, were not undergoing behavioral interventions for cognitive enhancement, demonstrated comprehension of verbal instructions, and had no limitations in basic activities of daily living. Exclusion criteria comprised individuals who (1) had been diagnosed with psychiatric disorders, (2) had been diagnosed with neurological conditions, and (3) expressed an aversion to computer usage.

### C. Intervention

The virtual shopping budget-management training was conducted twice a week for a total of 16 sessions over 8 weeks. The experimental group received the virtual shopping budget-management training program created with Scratch (MIT Media Lab, USA). This program presented four shopping lists with items located in two different aisles and clearly listed, eliminating the need for participants to memorize the items [15]. Participants were instructed to explore the virtual supermarket and purchase items from the shopping list. Each item had a different price, with lower prices yielding lower scores and higher prices resulting in higher scores. Participants were asked to maximize their scores within a predetermined budget while retaining knowledge of their scores. To make a payment, they were required to locate a cash desk, which then prompted the payment screen to appear. Following the payment, a statical screen with feedback on their performance appeared. This program places a particular emphasis on budget management within executive function training. It offers four levels of difficulty, challenging even healthy young adults, and utilizes various budget allocations to reduce learning effects.

### D. Measurement

Participant's executive function was assessed by measuring the activation of the dorsolateral prefrontal cortex (dlPFC) area using fNIRS while they performed the Trail Making Test Part B (TMT-B) and the Stroop test. To perform the TMT-B, participants were instructed to draw lines connecting numbers and letters alternately in ascending order [16]. The Stroop test typically comprises 2-4 performance tasks. In this study, participants engaged in the "interference color naming" task, where they named the color of words written in ink different from the word's meaning [17].

Activation in the dlPFC area was assessed through fNIRS (Octamon, Artinis, Netherlands), measuring hemodynamic responses within the dlPFC area. The fNIRS comprised a total of 8 channels, utilizing 760 nm and 850 nm infrared to detect changes in the cortical concentration of HbO2 and HHb [18, 19]. Participants wore this device while performing the TMT-B and the Stroop test. Hemodynamic response measurements continued until the completion of all assessments. Resting periods of indefinite duration were provided between each assessment until stability in activity was achieved. In this study, the mean values of HbO2 in the left and right dlPFC areas were utilized to measure hemodynamic responses. All data were sampled with a frequency of 10Hz with Oxysoft version 3.0.52.

### E. Statical analysis

All outcomes were analyzed using the SPSS 22.0 version (SPSS Inc., USA). To assess the normality of the sample, the Chi-square test was employed. A repeated two-way analysis of variance (ANOVA) was used to compare results between the two groups. Effects sizes were calculated using partial $\eta^2$ [20]. A statistical significance was set as $p < 0.05$.

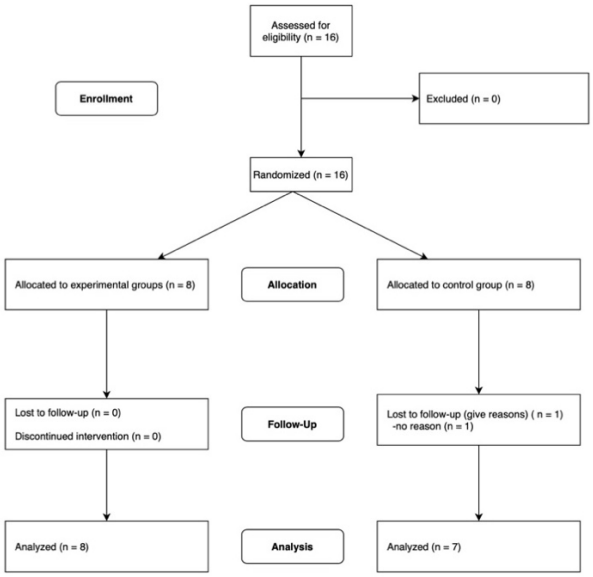



Fig. 1 Flowchart of the study

TABLE I

| Demographic characteristics | | Experimental group (n = 8) | Control group (n = 8) | $\chi^2$ |
|---|---|---|---|---|
| Sex | Male | 0 (0 %) | 3 (42.8 %) | .055 |
| | Female | 8 (100 %) | 4 (57.1 %) | |
| Age (years) | | 20.38 ± 0.92 | 21 ± 1.60 | .771 |

DEMOGRAPHIC CHARACTERISTICS OF THE GROUPS

## III. RESULTS

The flow of the study is illustrated in Figure 1. A total of 16 participants were randomly assigned, forming the experimental group with 8 participants and the control group with 8 participants. Following the pre-test, one participant in the control group dropped out. The results were analysed using data from 8 participants in the experimental group and 7 participants in the control group.

TABLE II
COMPARISON OF EXECUTIVE FUNCTION IN BOTH GROUPS

| Variables | | | Experimental group (n = 8) | Control group (n = 7) | Between-group differences (95% IC) | F | $\eta^2$ |
|---|---|---|---|---|---|---|---|
| TMT-B (sec) | | Pre-intervention | 19.01 ± 5.26 | 17.79 ± 6.75 | 1.3075 (-6.49; 3.87) | .297 | .022 |
| | | Post-intervention | 15.64 ± 4.88 | 15.73 ± 2.56 | | | |
| | | Within-group changes | -3.37 ± 2.13* | -2.06 ± 6.41 | | | |
| Stroop test (sec) | | Pre-intervention | 87.86 ± 18.94 | 90.43 ± 19.65 | 1.1294 (-11.59; 9.33) | .053 | .004 |
| | | Post-intervention | 77.62 ± 12.48 | 81.32 ± 16.32 | | | |
| | | Within-group changes | -10.24 ± 10.22* | -9.11 ± 8.52* | | | |
| fNIRS | TMT-B | Pre-intervention | 2.53 ± 1.12 | 2.75 ± 2.16 | -0.1876 (-2.20; 2.57) | .033 | .003 |
| | | Post-intervention | 2.30 ± 0.99 | 2.33 ± 1.76 | | | |
| | | Within-group changes | -.23 ± 1.55 | -.42 ± 2.43 | | | |
| | Stroop test | Pre-intervention | 2.62 ± 1.58 | 4.37 ± 4.33 | -1.2829 (-2.74; 5.31) | .569 | .042 |
| | | Post-intervention | 2.40 ± 1.25 | 2.47 ± 1.96 | | | |
| | | Within-group changes | -.21 ± 2.21 | -1.50 ± 4.21 | | | |

TMT-B: Trail Making Test Part B

Table 1 presents demographic characteristics, showing no significant differences between the two groups ($p > 0.05$) (Table 1).

Based on the results of the repeated two-way analysis of variance (ANOVA), there was no significant group x time interaction for the TMT-B ($p > 0.05$; $\eta^2$=.022) and the Stroop test ($p > 0.05$; $\eta^2$=.004). However, when comparing changes within each group, significant differences were found in the experimental group for both the TMT-B ($p < 0.05$; Change = -3.37) and the Stroop test ($p < 0.05$; Change = -10.24). In the control group, a significant difference was observed in the Stroop test ($p < 0.05$; Change = -9.11). The change value in the control group for TMT-B was -2.06.

There was no significant group x time interaction for HbO2 for the TMT-B ($p > 0.05$; $\eta^2$=.003) and the Stroop test ($p > 0.05$; $\eta^2$=.042) (Table 2). The change value in the experimental group for HbO2 for the TMT-B was -0.23, and for the Stroop test was -0.21. The change value in the control group for HbO2 for TMT-B was -0.42, and for the Stroop test was -1.5.

## IV. DISCUSSION

Executive function can commonly be categorized into three components: updating and monitoring of information ('updating'), inhibition of prepotent impulses ('inhibition'), and mental set shifting ('shifting) [21, 22]. Specifically, the ability to update information about prices, inhibit the impulse to buy, and think flexibly to select the most appropriate items within a budget is crucial for reasonable consumption within a budget [23], indicating the importance of executive function in shopping contexts. Previous research has revealed that individuals with executive function deficits, such as those with MCI or stroke, take longer to shop, exceed their budgets more frequently, and require more assistance compared to healthy individuals [7, 9]. This supports the notion that shopping could serve as an effective training method for enhancing executive function.

This study was designed to investigate the effects of virtual shopping budget-management training on executive function and related prefrontal cortex brain activity in healthy young adults. To achieve this, the experimental group participated in 16 training sessions, and when compared to the control group, no significant differences were observed in brain activity. This result contrasts with previous research findings [24].

One significant difference between the existing shopping training and the current study is our attempt to investigate the training's effects along with fNIRS, focusing on brain activation in the prefrontal cortex. Neuronal activation is crucial in neurorehabilitation, as it can lead to structural and functional changes in the brain through neuroplasticity [25, 26]. The brain area responsible for executive function is the dlPFC [27]. In this study, we measured HbO2 levels in this area and compared pre- and post-intervention differences. Although statistically significant results did not emerge, this effort is significant as the first attempt to demonstrate that virtual shopping training induces brain activation. But, there was no difference between the experimental and control groups. This lack of differentiation could be attributed to the assessment tool being too easy to evaluate healthy young adults, potentially resulting in a ceiling effect [28].

This study emphasizes the importance of advanced measurement techniques, such as neuroimaging, in evaluating cognitive training. Beyond statistical significance, it is essential to measure objective evidence, such as neural efficiency, to confirm intervention effects, surpassing the limitations of traditional assessments influenced by external factors like mood states [29, 30].

This study has several limitations. Firstly, the training was conducted on healthy young adults, despite the aim of enhancing executive function. Secondly, the limited sample size in this study hinders the generalization of the research findings. Thirdly, there were no significant results in HbO2 levels in the prefrontal cortex, which contrasts with previous research findings [24]. Individual differences in transfer outcomes are well-documented and it can be problematic, particularly with relatively small sample sizes typically gathered in brain imaging studies [31, 32]. Fourthly, there was

a difference in gender distribution between groups. According to previous studies, gender differences in executive function are not a significant factor, but there might be potential impacts in some aspects of executive function where gender differences are observed [33].

Future research should consider recruiting a larger sample, particularly with executive function impairments, for further analysis.

## V. CONCLUSION

In this study, the effects of virtual shopping budget-management training on brain activation in healthy young adults were not observed. Future research should include a larger sample size and appropriate participants.

## ACKNOWLEDGMENT

This research was supported by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (P0012724, The Competency Development Program for Industry Specialist).

## REFERENCES

[1] S. J. Gilbert and P. W. Burgess, "Executive function," *Current biology,* vol. 18, no. 3, pp. R110-R114, 2008.

[2] D. A. Cahn, E. V. Sullivan, P. K. Shear, A. Pfefferbaum, G. Heit, and G. Silverberg, "Differential contributions of cognitive and motor component processes to physical and instrumental activities of daily living in Parkinson's disease," *Archives of Clinical Neuropsychology,* vol. 13, no. 7, pp. 575-583, 1998.

[3] J.-H. Park, "Effects of cognitive-physical dual-task training on executive function and activity in the prefrontal cortex of older adults with mild cognitive impairment," *Brain & Neurorehabilitation,* vol. 14, no. 3, 2021.

[4] R. Kizony, N. Josman, N. Katz, D. Rand, and P. Weiss, "Virtual reality and the rehabilitation of executive functions: An annotated bibliography," *Isr. J of Occup. Ther,* vol. 17, no. 2, pp. E47-61, 2008.

[5] N. Josman and N. Katz, "Relationships of categorization on tests and daily tasks in patients with schizophrenia, post-stroke patients and healthy controls," *Psychiatry research,* vol. 141, no. 1, pp. 15-28, 2006.

[6] P. Mesa-Gresa, H. Gil-Gómez, J.-A. Lozano-Quilis, and J.-A. Gil-Gómez, "Effectiveness of virtual reality for children and adolescents with autism spectrum disorder: an evidence-based systematic review," *Sensors,* vol. 18, no. 8, p. 2486, 2018.

[7] S. Zygouris *et al.*, "Can a virtual reality cognitive training application fulfill a dual role? Using the virtual supermarket cognitive training application as a screening tool for mild cognitive impairment," *Journal of Alzheimer's Disease,* vol. 44, no. 4, pp. 1333-1347, 2015.

[8] J.-H. Park, "Does the virtual shopping training improve executive function and instrumental activities of daily living of patients with mild cognitive impairment?," *Asian Journal of Psychiatry,* vol. 69, p. 102977, 2022.

[9] S. Y. Nir-Hadad, P. L. Weiss, A. Waizman, N. Schwartz, and R. Kizony, "A virtual shopping task for the assessment of executive functions: Validity for people with stroke," *Neuropsychological rehabilitation,* vol. 27, no. 5, pp. 808-833, 2017.

[10] A. I. DREVER, E. ODDERS-WHITE, C. W. KALISH, N. M. ELSE-QUEST, E. M. HOAGLAND, and E. N. NELMS, "Foundations of Financial Well-Being: Insights into the Role of Executive Function, Financial Socialization, and Experience-Based Learning in Childhood and Youth," *Journal of Consumer Affairs,* vol. 49, no. 1, pp. 13-38, 2015.

[11] M. Jacoby, S. Averbuch, Y. Sacher, N. Katz, P. L. Weiss, and R. Kizony, "Effectiveness of executive functions training within a virtual supermarket for adults with traumatic brain injury: a pilot study," *IEEE transactions on neural systems and rehabilitation engineering,* vol. 21, no. 2, pp. 182-190, 2013.

[12] Y. Hoshi and M. Tamura, "Near-infrared optical detection of sequential brain activation in the prefrontal cortex during mental tasks," *NeuroImage,* vol. 5, no. 4, pp. 292-297, 1997.

[13] S. Tak and J. C. Ye, "Statistical analysis of fNIRS data: a comprehensive review," *Neuroimage,* vol. 85, pp. 72-91, 2014.

[14] P. Pinti, I. Tachtsidis, A. Hamilton, J. Hirsch, C. Aichelburg, S. Gilbert, and P. W. Burgess, "The present and future use of functional near‐infrared spectroscopy (fNIRS) for cognitive neuroscience," *Annals of the New York Academy of Sciences,* vol. 1464, no. 1, pp. 5-29, 2020.

[15] D. Rand, N. Katz, and P. L. Weiss, "Evaluation of virtual shopping in the VMall: Comparison of post-stroke participants to healthy control groups," *Disability and rehabilitation,* vol. 29, no. 22, pp. 1710-1719, 2007.

[16] E. Zayat, M. Rempfer, B. Gajewski, and C. E. Brown, "Patterns of association between performance in a natural environment and measures of executive function in people with schizophrenia," *Psychiatry research,* vol. 187, no. 1-2, pp. 1-5, 2011.

[17] T. Y. Kim *et al.*, "Development of the Korean Stroop Test and Study of the Validity and the Reliability," *Journal of the Korean Geriatrics Society,* vol. 8, no. 4, pp. 233-240, 2004.

[18] V. de Belli *et al.*, "Prefrontal cortical activity during preferred and fast walking in young and older adults: An fNIRS Study," *Neuroscience,* vol. 473, pp. 81-89, 2021.

[19] P. Nóbrega-Sousa, L. T. B. Gobbi, D. Orcioli-Silva, N. R. d. Conceição, V. S. Beretta, and R. Vitório, "Prefrontal cortex activity during walking: effects of aging and associations with gait and executive function," *Neurorehabilitation and Neural Repair,* vol. 34, no. 10, pp. 915-924, 2020.

[20] J. Cohen, "Statistical power analysis for the behavioral sciences. Lawrence Erlbaum," *Hillsdale, NJ,* pp. 75-108, 1988.

[21] R. F. Baumeister, B. J. Schmeichel, and K. D. Vohs, "Self-regulation and the executive function: The self as controlling agent," *Social psychology: Handbook of basic principles,* vol. 2, pp. 516-539, 2007.

[22] A. Miyake, N. P. Friedman, M. J. Emerson, A. H. Witzki, A. Howerter, and T. D. Wager, "The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis," *Cognitive psychology,* vol. 41, no. 1, pp. 49-100, 2000.

[23] R. F. Baumeister, E. A. Sparks, T. F. Stillman, and K. D. Vohs, "Free will in consumer behavior: Self-control, ego depletion, and choice," *Journal of Consumer Psychology,* vol. 18, no. 1, pp. 4-13, 2008.

[24] L. Nguyen, K. Murphy, and G. Andrews, "Cognitive and neural plasticity in old age: A systematic review of evidence from executive functions cognitive training," *Ageing research reviews,* vol. 53, p. 100912, 2019.

[25] V. Demarin and S. MOROVIĆ, "Neuroplasticity," *Periodicum biologorum,* vol. 116, no. 2, pp. 209-211, 2014.

[26] M. Puderbaugh and P. D. Emmady, "Neuroplasticity," in *StatPearls [Internet]*: StatPearls Publishing, 2023.

[27] A. Nieder, "Prefrontal cortex and the evolution of symbolic reference," *Current opinion in neurobiology,* vol. 19, no. 1, pp. 99-108, 2009.

[28] E. Judson, "Learning about bones at a science museum: examining the alternate hypotheses of ceiling effect and prior knowledge," *Instructional Science,* vol. 40, pp. 957-973, 2012.

[29] L. G. Chepenik, L. A. Cornew, and M. J. Farah, "The influence of sad mood on cognition," *Emotion,* vol. 7, no. 4, p. 802, 2007.

[30] Y. J. Jeun, Y. Nam, S. A. Lee, and J.-H. Park, "Effects of Personalized Cognitive Training with the Machine Learning Algorithm on Neural Efficiency in Healthy Younger Adults," *International Journal of Environmental Research and Public Health,* vol. 19, no. 20, p. 13044, 2022.

[31] P. Reddy, P. A. Shewokis, and K. Izzetoglu, "Individual differences in skill acquisition and transfer assessed by dual task training performance and brain activity," *Brain Informatics,* vol. 9, no. 1, p. 9, 2022.

[32] J. Salmi *et al.*, "Working memory training restores aberrant brain activity in adult attention-deficit hyperactivity disorder," *Human Brain Mapping,* vol. 41, no. 17, pp. 4876-4891, 2020.

[33] N. M. Grissom and T. M. Reyes, "Let's call the whole thing off: evaluating gender and sex differences in executive function," (in eng), *Neuropsychopharmacology,* vol. 44, no. 1, pp. 86-96, Jan 2019.

# Machine-Learning Based Analysis of Fall Risk in Hospitalized Patients Using Electronic Medical Records and Gait Data Captured Through IMU Sensors

Jung-Yeon Kim[1,*], Chomyong Kim[2], Seob Jeon[2], Hyo-Wook Gil[2], Ik-Dong Yu[2], Ji-Won Lyu[2], Euy-Hyun Chung[2], and Yunyoung Nam[3]

[1]*ICT Convergence Research Center, Soonchunhyang University, Asan, South Korea*

[2]*Soonchunhyang University Cheonan Hospital, Cheonan, South Korea*

[3]*Department of Computer Science and Engineering, Soonchunhyang University, Asan, South Korea*

*Contact: betterwitme@sch.ac.kr

*Abstract*— **Falling is one of major causes of death and closely associated with quality of living in older adults. Although investigation of fall detection methods has been an active research field and effective methods have been proposed, relatively less attention has drawn to the field of identification of older adults at high risk of fall relatively recently due to the complex nature of fall risk analysis and it is understood that there are rooms for improvement. In this study, we aimed to identify inpatients at high risk of falling by applying machine-learning models on electronic medical records (EMRs) and dataset on gait collected from 42 hospitalized patients. The result indicates that identification of hospitalized patients at high risk of falling can be achieved by utilizing EMRs, and comparison of classification performance suggests that information on daily gait performance can be beneficial for the models to better identify the patients at high risk of falling.**

## I. INTRODUCTION

As people are aging, muscle strength weakens that may be associated with increase in fall incidents among adults aged 65 and over. Fall is known to be one of major cause of death and affects the quality of life due to reduced mobility. It has been reported that most common adverse events reported in hospitals are patient falls indicating that likelihood of falls significantly increase [1]. Owing to this, it has been an active research topic to analyze risk factors of falls in older adults [2]. Electronic medical records (EMRs) have been a popular data source in studies where aim to identify predictors that can be used for fall predictive models [3]. In addition, there are significant numbers of studies investigating applying advanced sensor technologies for collecting clinically meaningful data and machine learning techniques to analyze them in the field [4-8]. However, majority of the studies seem to focus on short term sensing data collected during assessing patients using clinical measures for gait or balance functions [4, 9]. Advancements in sensor technology have made sensors more suitable for long-term monitoring as they can be worn easily and operating for days without having to change batteries. Long-lasting sensors can be beneficial in terms of capturing pathological characteristics that may not be seen during clinical examinations. Although analyzing long-term monitoring data can be problematic as long-term monitoring data are high in volumes due to the relatively high sampling rate and they are usually noisy due to uncontrolled data collection environment. However, clinically meaningful information can be obtained with the help of appropriate signal processing methods and machine-learning techniques.

In this study, we aimed to investigate if patients at risk of fall can be identified using EMRs and gait performance parameters derived from long-term monitoring data collected through IMU sensors placed on both ankles in hospitalized patients. The detailed information on the proposed method and validation results of the method are as follows.

## II. METHODS

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it or use the paste function.

### *A. Proposed method for fall risk classification*

The proposed method utilized electronic medical records (EMRs) and gait parameters to categorize hospitalized patients to different risk groups. Machine learning techniques were used to implement two functions. Firstly, the techniques were used to identify gait patterns from IMU data which records information related to movements, such as angular velocity as well as acceleration. The recognized gait patterns were analyzed to obtain gait parameters that reflect gait performance. Next, the models were trained using EMRs and gait parameters to classify the patients into groups with different levels of fall risk. Detailed information on variables of EMRs and gait parameters obtained from IMU dataset are described in the section of model predictors.

### *B. Data collection*

Hospitalized patients (n=42) were recruited for data collection. The data include EMRs and IMU signals. The former collects daily clinical treatment information while the latter collects daily movement related data collected from IMU sensors placed on both ankles of the patients. Informed written

consent was obtained prior to enrolment. For IMU data, the patients were asked to wear the same type of IMU sensor on their both ankles. They were only required to wear the device and do not change the location or direction of the placement. No other instructions were given to the patients during the data collection process.

### C. Model predictors

Variables of EMRs investigated in this study includes patients' characteristics, results of assessments related to ulcer, activities of daily living (ADLs), and Morse Fall Scale (MFS) result. When patients are admitted to the hospital, degree of fall risk is assessed by nurses using MFS, and the patients were labelled according to this score rated when admitted. In practice, MFS score can be used to categorize patients into three fall risk groups: low, moderate, and high risk: low (0~24), moderate (25~44), and high risk (45+) [10]. According to MFS classification criteria, the hospitalized patients recruited for this study were categorized into the low-, moderate-, and high-risk groups were 6, 14, and 22, respectively. On the other hand, gait related parameters were obtained from IMU signals. Firstly, gait pattern signals were identified from the total IMU data using machine-learning. How much time a patient was walking while wearing the IMU sensors can be estimated during this process, and this information was used as one of features for hospitalized patients at high risk. From the part of IMU signals that reflect gait patterns were analyzed and gait parameters were extracted from both feet. The ratio of each feature of left foot to right foot was considered as a feature representing gait performance affected by both feet. Totally, 21 features were extracted from the IMU signals that contain gait patterns. Both types of feature sets are summarized in Table I.

TABLE I
FEATURES USED TO IDENTIFY PATIENTS AT RISK OF FALL

| Feature Source (N) | Features |
|---|---|
| EMRs (18) | Sex, age, weight, ADL-defecation, ADL-repositioning, ADL-bed-exit, ADL-eating, ADL-danger, ADL-professional, Ulcer-behavior, Ulcer-nutrition, Ulcer-mobility, use of breathing support device, pain level, use of assistive device, activity level, level of consciousness, fall history |
| IMUs (21) | Mean, STD, VAR, IQR, Median, of swing, stance, and stride time, dominant frequency, power, energy, percentage of being in active state |

### D. Model validation

In this study, nine classifiers were tested to determine the degree of fall risk in hospitalized patients including k-nearest neighbor (KNN), gradient boosting classifier (GBC), logistic regression (LR), random forest (RF), support vector machine (SVM), multi-layer perceptron (MLP), AdaBoost (ADA), XGBoost (XGB), and a classifier obtained by applying ensemble learning. K-fold (k=10) validation was employed to test model performance. Features were scaled to zero mean and unit variance with respect to the training dataset. In addition, datasets were resampled by applying the synthetic minority oversampling technique to overcome imbalance between classes.

## III. RESULTS

The degree of fall risk was initially classified into two degrees: low and high risk, and it was extended to three classes: low-, moderate-, and high-risk. The results obtained under two different classification scheme follows as below.

### A. Classification of patients at risk of fall into two degrees: low and high risk

Classification performance of hospitalized patients to low- and high-risk groups are summarized in Table II. Classification result using the EMR feature set ranges from 0.50 to 0.88, and mean accuracy was 0.70 ± 0.26. There was an increase in mean accuracy by 0.01 when the IMU feature set used. The highest mean accuracy (0.87) was obtained when both EMR and IMU feature sets were used for classification. Not only the performance was increased, but also the trained models were relatively more reliable according to standard deviation value (±0.21). Comparison of performance between model revealed that RF (0. 96 ± 0.14) outperforms other classifiers, and it was followed by ADA and GBC. It is interesting to note that the results achieved by applying ensemble learning did not improve the performance significantly.

TABLE II
CLASSIFICATION OF RISK INTO TWO DEGREES: LOW AND HIGH RISK

| Model | Mean Accuracy (Standard Deviation) | | |
|---|---|---|---|
| | EMR | IMU | EMR + IMU |
| KNN | 0.72 ± 0.30 | 0.75 ± 0.25 | 0.69 ± 0.24 |
| GBC | 0.84 ± 0.31 | 0.84 ± 0.26 | 0.90 ± 0.20 |
| LR | 0.82 ± 0.29 | 0.81 ± 0.24 | 0.84 ± 0.26 |
| RF | 0.93 ± 0.18 | 0.91 ± 0.23 | 0.96 ± 0.14 |
| SVM | 0.76 ± 0.30 | 0.76 ± 0.25 | 0.81 ± 0.27 |
| MLP | 0.50 ± 0.00 | 0.94 ± 0.16 | 0.88 ± 0.24 |
| ADA | 0.81 ± 0.30 | 0.87 ± 0.25 | 0.91 ± 0.19 |
| XGB | 0.81 ± 0.34 | 0.84 ± 0.26 | 0.90 ± 0.20 |
| Ensemble | 0.88 ± 0.30 | 0.90 ± 0.24 | 0.94 ± 0.16 |
| Mean | 0.79 ± 0.26 | 0.85 ± 0.23 | 0.87 ± 0.21 |

Table III shows the order of top 10 features according to feature importance obtained by top 3 models trained with both EMR and IMU feature sets. Interestingly, the percentage of being active was ranked firstly by every model. In addition, similar number of EMR and IMU features were ranked within 10 variables among 39 features, except for ADA model as most of the top 10 ranked features were derived from IMU dataset apart from age and weight. In terms of gait performance parameters, RF and GBC consider features related to swing phase more important than stance phases as the number of such feature type were greater than stance phase or stride. On the other hand, ADA model take relatively similar interests on features related to swing and stance phases, and stride. Lastly, weight and age seem to play an important role as they were ranked within top 10 features for the top 3 models that determined the levels of fall risk into two classes.

TABLE III
FEATURE IMPORTANCE ACCORDING TO TOP 3 MODELS TRAINED WITH BOTH EMR AND IMU DATASETS

| Rank | Top 3 Models | | |
|---|---|---|---|
| | **RF** | **GBC** | **ADA** |
| 1 | Percentage of being active | Percentage of being active | Percentage of being active |
| 2 | Power | Swing phase ratio | Age |
| 3 | Median stride time | Weight | 1-step cycle |
| 4 | Energy | Energy | Weight |
| 5 | STD of swing phase | Power | IQR of stride |
| 6 | IQR of swing phase | Age | Median of swing phase |
| 7 | Weight | Median swing phase | Power |
| 8 | Use of assistive device | VAR of swing phase | Mean stride time |
| 9 | Age | Ulcer-nutrition | Median stride time |
| 10 | Ulcer-nutrition | STD of stride | Mean stance phase |

### *B. Classification of patients at risk of fall into three degrees: low, moderate, and high risk*

Table IV summarizes classification performance of hospitalized patients to low-, moderate-, and high-risk groups. Classification accuracy obtained by using EMR feature set ranged from 0.61 to 0.74, and mean accuracy was 0.70 ± 0.26. There was slight increase in the mean accuracy by 0.01 when IMU feature set used. It increased by 0.05 when trained with both EMR and IMU feature sets. Moreover, trained models were also relatively more reliable according to standard deviation (±0.21). There was an inverse relation between standard deviation and the number of features, and the lowest standard deviation was found when both EMR and IMU feature sets were used. The highest mean accuracy (0.80) was obtained with the ensemble model when both feature sets were used for classification. Model comparison shows similar result that the performance of RF was better than other classifiers, followed by the ensemble model, ADA, and XGBoost. The mean accuracy of the rest classifiers was less than 0.90.

TABLE IV
CLASSIFICATION OF RISK INTO THREE DEGREES: LOW, MODERATE, AND HIGH RISK

| Model | Mean Accuracy (Standard Deviation) | | |
|---|---|---|---|
| | **EMR** | **IMU** | **EMR + IMU** |
| **KNN** | 0.61 ± 0.28 | 0.62 ± 0.23 | 0.65 ± 0.24 |
| **GBC** | 0.73 ± 0.24 | 0.70 ± 0.20 | 0.83 ± 0.19 |
| **LR** | 0.64 ± 0.26 | 0.68 ± 0.26 | 0.71 ± 0.25 |
| **RF** | 0.74 ± 0.24 | 0.77 ± 0.18 | 0.82 ± 0.17 |
| **SVM** | 0.70 ± 0.30 | 0.61 ± 0.28 | 0.71 ± 0.23 |
| **MLP** | 0.67 ± 0.22 | 0.79 ± 0.21 | 0.67 ± 0.20 |
| **ADA** | 0.67 ± 0.32 | 0.64 ± 0.22 | 0.73 ± 0.26 |
| **XGB** | 0.73 ± 0.24 | 0.68 ± 0.27 | 0.83 ± 0.22 |
| **Ensemble** | 0.74 ± 0.24 | 0.80 ± 0.22 | 0.85 ± 0.19 |
| **Mean** | 0.70 ± 0.26 | 0.71 ± 0.23 | 0.76 ± 0.21 |

Top 10 features are listed in Table V determined by RF, GBC, and XGB trained to classify three different risk groups. Percentage of being active was ranked firstly by RF and GBC models, and weight and age were ranked within top 10 features, except for XGB, in which the percentage of being active was not considered important and only age was ranked within top 10 features. Unlike the result of feature importance for classifying two classes, features derived from IMU data were found to be more sensitive as the number of features from the dataset slightly increased.

TABLE V
FEATURE IMPORTANCE ACCORDING TO TOP 3 MODELS TRAINED WITH BOTH EMR AND IMU DATASETS

| Rank | Top 3 Models | | |
|---|---|---|---|
| | **RF** | **GBC** | **XGB** |
| 1 | Percentage of being active | Percentage of being active | ADL-defecation |
| 2 | IQR of swing phase | Age | Dominant frequency |
| 3 | Swing phase ratio | Dominant frequency | Power |
| 4 | Age | Stance phase ratio | Stance phase ratio |
| 5 | Weight | Swing phase ratio | Percentage of being active |
| 6 | Stance phase ratio | IQR of swing phase | Age |
| 7 | Mean stance phase | ADL-defecation | Mean swing phase |
| 8 | Dominant frequency | Energy | Breathing support |
| 9 | Energy | Weight | IQR of swing phase |
| 10 | Median swing phase | Mean swing phase | Mean stance phase |

Table VI shows the confusion matrix of classification result indicating that moderate-risk group can be effectively identified as there was only a case that was misclassified to high-risk group. On the other hand, the trained models misclassified 5 cases from high- to low-risk groups. In addition, the models were having troubles to correctly identify low-risk groups.

TABLE VI
CONFUSION MATRIX OF CLASSIFICATION RESULT OBTAINED FROM THE RANDOM FOREST MODEL FOR THREE CLASSES USING BOTH EMR AND IMU DATASETS

| True | Predicted | | |
|---|---|---|---|
| | **Low** | **Moderate** | **High** |
| **Low** | 16 | 4 | 2 |
| **Moderate** | 0 | 21 | 1 |
| **High** | 5 | 0 | 17 |

## IV. DISCUSSION AND CONCLUSIONS

This study reported the proposed method that aims to classify patients at high risk of fall using patient information acquired when admitted and updated periodically and gait performance monitored daily where the data collected from wearable sensors attached to both ankles while hospitalized. Totally, nine machine learning classifiers were tested. The result indicates

that identification of patient at high risk of fall can be achieved accurately as the classification performance reached up to 96% when both variables of EMRs and gait parameters were fed to the tested models. In addition, combining both types of features helps the models to better perform classification while keeping the model more stable as the standard deviation values were relatively less than the model trained with each feature set alone. On the contrary, categorizing the hospitalized patients to either low-, moderate-, or high-risk group was less satisfactory as the models only successfully identified the moderate-risk groups while poorer performance was shown for the classification of low- and high-risk groups as the majority misclassifications were classifying the patients in the low-risk group to the high-risk group or vice versa. This finding can be interpreted as either MFS cutoff score used to categorize three risk groups in this study may require adjustment to better separate the patients into more suitable risk groups or using only MFS score to categorize risk levels into three degrees may be limited that additional measures can support the MFS. Although the study findings are limited to some extent, the proposed method showed high classification performance for identifying fall risk degrees by utilizing daily hospital care records and gait performance.

## ACKNOWLEDGMENT

This research was supported by Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE). (P0012724, The Competency Development Program for Industry Specialist) .

## REFERENCES

[1] I. M. Miake-Lye, S. Hempel, D. A. Ganz, and P. G. Shekelle, "Inpatient Fall Prevention Programs as a Patient Safety Strategy," *Annals of Internal Medicine,* vol. 158, no. 5_Part_2, pp. 390-396, 2013/03/05, 2013.

[2] H. Nagano, and R. K. Begg, "Shoe-Insole Technology for Injury Prevention in Walking," *Sensors,* vol. 18, no. 5, pp. 1468, 2018.

[3] D. S. Lindberg, M. Prosperi, R. I. Bjarnadottir, J. Thomas, M. Crane, Z. Chen, K. Shear, L. M. Solberg, U. A. Snigurska, Y. Wu, Y. Xia, and R. J. Lucero, "Identification of important factors in an inpatient fall risk prediction model to improve the quality of care using EHR and electronic administrative data: A machine-learning approach," *International Journal of Medical Informatics,* vol. 143, pp. 104272, 2020/11/01/, 2020.

[4] S. K. Byrnes, C. Nüesch, S. Loske, A. Leuenberger, S. Schären, C. Netzer, and A. Mündermann, "Inertial Sensor-Based Gait and Attractor Analysis as Clinical Measurement Tool: Functionality and Sensitivity in Healthy Subjects and Patients With Symptomatic Lumbar Spinal Stenosis," *Frontiers in Physiology,* vol. 9, 2018-August-14, 2018.

[5] D. Rodríguez-Martín, C. Pérez-López, A. Samà, J. Cabestany, and A. Català, "A wearable inertial measurement unit for long-term monitoring in the dependency care area," *Sensors,* vol. 13, no. 10, pp. 14079-14104, 2013.

[6] A. Mehmood, M. A. Khan, M. Sharif, S. A. Khan, M. Shaheen, T. Saba, N. Riaz, and I. Ashraf, "Prosperous Human Gait Recognition: an end-to-end system based on pre-trained CNN features selection," *Multimedia Tools and Applications*, 2020/04/28, 2020.

[7] C. Tunca, G. Salur, and C. Ersoy, "Deep learning for fall risk assessment with inertial sensors: Utilizing domain knowledge in spatio-temporal gait parameters," *IEEE journal of biomedical and health informatics,* vol. 24, no. 7, pp. 1994-2005, 2019.

[8] A. Sampath Dakshina Murthy, T. Karthikeyan, and R. Vinoth Kanna, "Gait-based person fall prediction using deep learning approach," *Soft Computing*, pp. 1-9, 2021.

[9] J. Y. Kim, S. Lee, H. B. Lee, B.-G. Kang, S.-B. Im, and Y. Nam, "Gait analysis in patients with neurological disorders using ankle-worn accelerometers," *The Journal of Supercomputing,* vol. 77, no. 8, pp. 8374-8390, 2021/08/01, 2021.

[10] J. M. Morse, C. Black, K. Oberle, and P. Donahue, "A prospective study to identify the fall-prone patient," *Social Science & Medicine,* vol. 28, no. 1, pp. 81-86, 1989/01/01/, 1989.

# Falling Direction Prediction Using Artificial Intelligence: Utilizing Body Keypoint Coordinates as Feature Extraction in Falling Direction Prediction

Chomyong Kim[1], Jung-Yeon Kim[1], Yunyoung Nam[2*]
[1]ICT Convergence Research Centre, Soonchunhyang University, Asan, South Korea
[2]Department of Computer Science and Engineering, Soonchunhyang University, Asan, South Korea
*Contact: ynam@sch.ac.kr

*Abstract*— **Predicting falling directions is crucial for anticipating potential hazards and enabling timely preventive measures, minimizing the risks associated with falls. This proactive approach enhances safety and minimizes the impact of fall-related incidents. However, most recent studies have been limited to classifying the presence of falls, and research on predicting the direction of falls is scarce and hard to find.**

**This study introduces an AI-powered method for classifying falling directions. By harnessing advanced AI techniques, we establish the viability of accurately determining fall directions in diverse scenarios. Our approach not only enhances safety protocols but also enables rapid responses during falls by predicting their directions. Thorough analysis and model training underpin the reliability of our AI-based system, laying a strong foundation for its practical implementation. Through comprehensive analysis and model training, our findings establish a foundation for the effective implementation of AI-driven falling direction classification systems.**

## I. Introduction

Predicting the direction of falling is an important topic for the health and safety of the elderly. Falls are a common accident among the elderly, which can lead to fractures, bruises, and serious injuries. Predicting the direction of falling can help identify the cause of falls and prevent them, thereby protecting the health and safety of the elderly. Falls can cause various injuries to the elderly, which can lead to restrictions in daily life. Predicting the direction of falling helps identify situations where the elderly are at high risk of falling and enables them to take preventive measures. Predicting the direction of falling also helps medical professionals and caregivers evaluate the health status of the elderly and provide appropriate management methods. This can improve the health status of the elderly and minimize injuries caused by falls [1].

An uncontrolled fall can escalate into even more dire situations than the initial cause itself, especially if the unattended patient is not promptly transported to a medical facility. To mitigate the adverse outcomes of such unfortunate incidents, there has been a substantial increase in the demand for intelligent systems designed to prevent, detect, and report these occurrences over the past decade. Up to this point, numerous studies have been proposed, addressing various facets of the fall detection issue, ranging from straightforward practical systems to intricate ones involving detection algorithms and methods for feature extraction [2].

## II. Related Work

In the field of fall behavior classification research, the primary focus is on classifying the presence or absence of falls and classifying the direction of falls. Most of the research in this field heavily relies on sensors or image-based studies.

Using sensors, in the study by Dai et al. (2010) [3] and various other articles, falls are categorized into three types: forward falls, backward falls, and lateral falls. Meanwhile, the research conducted by Jian and Chen (2015) [4] specifically identifies backward falls. In the work of Ando et al. (2016) [5], in addition to detecting falls, several everyday activities are also recognized, including walking, sitting, running, climbing stairs, and lying down.

There are numerous studies that utilize images for classifying the presence of falls; however, research for predicting the direction is limited. Furthermore, although there are open-source datasets available for falls, including various Activities of Daily Life (ADL), most of them consist of falls without specific ADL categorization. For studies focusing on fall direction classification, the 'UP-Fall Detection Dataset,' introduced by Martinez et al. (2019) [7], is available. This dataset is designed for multimodal research and captures data using sensors, RGB cameras, and Depth cameras. It includes ADL activities such as Walking, Standing, Jumping, Sitting, Laying, Picking up an object, as well as falls categorized into three directions: Falling backward, Falling forward using hands, Falling forward using knees, Falling sideward, and Falling while sitting in an empty chair.

Sensor data can yield better results than images in confirming the presence of falls and classifying their direction in research [8]. However, it can be quite challenging for individuals to consistently attach sensors in the same position every time. However, in the case of cameras, as they are not attached to the body, they are expected to be more practical from a usability standpoint.

## III. Data Description

This dataset consists of paired video-sensor data, utilizing 8 cameras and 12 sensors. Various scenarios involving falls and

ADL were set up and then recorded in locations such as hospitals, nursing homes, homes, and outdoor environments. The types of falls are categorized based on their direction, including forward falls, lateral falls, and backward falls. The term "non-fall" inherently signifies the absence of falling. Nevertheless, in the context of this dataset, instances where actions closely resembled the precipice of falling were also classified as non-falls. Additionally, it is noteworthy that the dataset employed for research purposes deliberately omitted scenarios depicting routine or normal behaviors.


Fig. 1, 2 Camera placement, Sensor wearing position

### A. Forward Fall (Anterior Fall)

A forward fall signifies falling forward, where the body leans forward and hits the ground. This type of fall is characterized by the front of the body coming into contact with the ground. Forward falls pose a higher risk of injuries to the face, chest, abdomen, hands, knees, and other frontal areas. An example of a forward fall is when someone slips and falls forward while walking.

### B. Backward Fall (Posterior Fall)

A backward fall refers to falling backward, where the body tilts backward and the back or spine makes contact with the ground. Backward falls are more likely to result in injuries to the head, spine, pelvis, hips, elbows, wrists, and other rearward areas. An example of a backward fall is slipping and falling backward or stumbling backward.

### C. Lateral fall (Sideways Fall)

A lateral fall indicates falling to the side, with the body tilting to one side, and either the side or lateral part of the body hitting the ground. Lateral falls often lead to injuries on the side, including the pelvis, hips, elbows, shoulders, knees, and more. An example of a lateral fall is when one leg slips, causing a person to fall to the side.

The video footage captured by 8 cameras was synchronized with the sensor data, and 10-second segments were extracted, including pre-fall, fall, post-fall, and recovery phases. The video was recorded at 60 frames per second (fps), while the sensor data was collected at 60Hz.

## IV. Experiments and Results

The objective of this study is to classify falls and non-falls and, furthermore, predict the type of falls, i.e., their direction. The dataset used in this study was recorded in a hospital room and targeted adult males and females. This study utilized a total of 47 videos each for forward falls, backward falls, lateral falls, and non-falls, recorded over two days. The purpose was to perform keypoint extraction to utilize the movement of coordinate values. To increase the number of videos for analysis, all videos were combined into their respective classes without categorizing them by camera, and they were utilized accordingly.

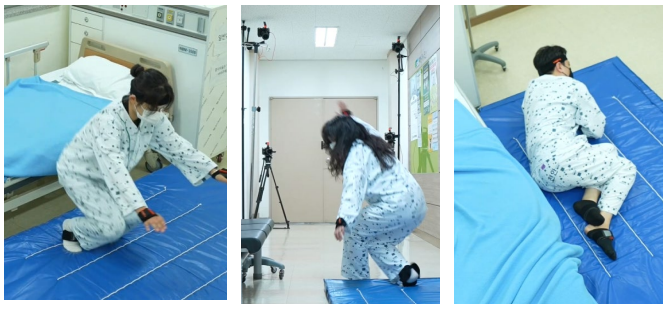


Fig. 3 Example of Dataset (Forward, Backward, Lateral)

### A. Feature Extraction

Keypoints hold significance in the domain of human pose classification due to their ability to encapsulate the spatial interconnections among various body segments. The extraction of keypoints enables us to depict the human pose as a collection of spatial points, thereby offering a suitable input for machine learning algorithms employed in the classification process [9].

The Keypoint Extraction feature within Mediapipe primarily serves the purpose of analyzing human body movements and facilitating pose tracking. Mediapipe's Keypoint Extraction process entails the detection and retrieval of key anatomical points from real-time video streams or images, representing various aspects of the human body. These key points encompass critical landmarks like joints, facial features, and fingertip positions. The Mediapipe model provides coordinates that represent the positions of detected keypoints in the input images or video. These coordinates correspond to various anatomical structures, depending on the specific type of keypoints being analyzed. In this study, only the coordinate values were utilized, and the landmarks themselves were not drawn.

### B. Machine Learning-based Classifier

Traditional machine learning methods such as Decision Tree, Random Forest, Stochastic Gradient Descent (SDG), and K-Nearest Neighbors (KNN) were employed.

| Model | Class | Precision | Recall | F1-Score |
|---|---|---|---|---|
| **Decision Tree** | Forward | 0.56 | 0.67 | 0.61 |
| | Backward | 0.68 | 0.61 | 0.64 |
| | Sideways | 0.61 | 0.63 | 0.62 |
| | Non-Fall | 0.94 | 0.86 | 0.90 |
| | ACC | 0.708 | | |
| **Random Forest** | Forward | 0.78 | 0.65 | 0.71 |
| | Backward | 0.62 | 0.89 | 0.73 |
| | Sideways | 0.76 | 0.62 | 0.68 |
| | Non-Fall | 0.76 | 0.62 | 0.68 |
| | ACC | **0.77** | | |
| | Forward | 0.72 | 0.56 | 0.63 |

| | | | | |
|---|---|---|---|---|
| **SGD Classifier** | Backward | 0.65 | 0.65 | 0.65 |
| | Sideways | 0.60 | 0.68 | 0.63 |
| | Non-Fall | 0.92 | 0.94 | 0.93 |
| | ACC | 0.735 | | |
| **KNN Classifier** | Forward | 0.47 | 0.82 | 0.60 |
| | Backward | 0.53 | 0.62 | 0.57 |
| | Sideways | 0.61 | 0.16 | 0.26 |
| | Non-Fall | 0.71 | 0.71 | 0.71 |
| | ACC | 0.572 | | |

Table 1. Result of Machine Learning-based Classifier

### *C. Deep Learning-based Classifier*

1) LSTM

The superior performance of LSTM (Long Short-Term Memory) compared to traditional algorithms like SVM (Support Vector Machine) and ANN (Artificial Neural Network) can be attributed to its capacity to effectively leverage long-term memory for the prediction of text sequences [9]. LSTM is primarily employed in text analysis, but due to its effectiveness in predicting sequences, it is also commonly used in models for classifying movements. Several studies utilize body keypoints values in this regard. While LSTM models are known for their strength in handling sequences and have been employed in research such as Human Pose Estimation (HPE), in this particular study, when classifying three different fall directions and non-falls, the accuracy achieved was 27%. Similarly, for the classification of forward, backward, and lateral falls, the accuracy was observed to be 32%.

2) VGG16

In addition to utilizing keypoints for analysis, we also conducted a straightforward VGG-based study. However, to account for potential variations in values due to the camera's position, we employed video footage captured from just one of the eight available camera angles. Notably, we refrained from conducting separate feature extraction and directly trained the model. The outcome of this approach yielded an accuracy of 27%.

## V. Conclusions & Limitation

This study began with the hypothesis that it is possible to predict the direction of falls based on the coordinates extracted from body keypoints in videos recorded from eight different angles. The research focused on using coordinate values without visualizing them, aiming to predict the direction of falls in response to abrupt coordinate movements. As a result, traditional machine learning classifiers outperformed the robust LSTM model for sequence data. Particularly noteworthy was the Random Forest model, which achieved an accuracy of 77% without undergoing separate hyperparameter tuning, making it the most proficient. It can be inferred that classifying forward and lateral falls is more challenging compared to posterior falls. This could suggest that there isn't a significant difference between forward falls and lateral falls. Furthermore, the inefficiency of LSTM may be attributed to the brief duration of falls, approximately 0.5 seconds, during which the necessary feature extraction for training may not have been effectively carried out. Therefore, future research endeavors will involve the utilization of open-source alternatives, excluding Mediapipe, to extract body keypoints. Additionally, feature extraction through transfer learning will be performed for each frame, followed by V incorporates the efficient machine learning classifiers observed in this study. This study holds value as foundational research that utilizes images and artificial intelligence to predict and develop the direction of object movement.

## Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ICAN(ICT Challenge and Advanced Network of HRD) program(IITP-2023-2020-0-01832) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation)

## References

[1] Merrill R. Landers, Sarrie Oscar, Jessica Sasaoka, Kyle Vaughn, Balance Confidence and Fear of Falling Avoidance Behavior Are Most Predictive of Falling in Older Adults: Prospective Analysis, Physical Therapy, Volume 96, Issue 4, 1 April 2016, Pages 433–442, https://doi.org/10.2522/ptj.20150184

[2] Jahanjoo, A., Naderan, M. & Rashti, M.J. Detection and multi-class classification of falling in elderly people by deep belief network algorithms. J Ambient Intell Human Comput 11, 4145–4165 (2020). https://doi.org/10.1007/s12652-020-01690-z

[3] Dai J, Bai X, Yang Z, Shen Z, Xuan D (2010) Mobile phone-based pervasive fall detection. Pers Ubiquit Comput 14(7):633–643

[4] Jian H, Chen H (2015) A portable fall detection and alerting system based on k-NN algorithm and remote medicine. China Commun 12(4):23–31

[5] Ando B, Baglio S, Lombardo C, Marletta V (2016) A multisensor data-fusion approach for ADL and fall classification. IEEE Trans Instrum Meas 65(9):1960–1967

[6] Martínez-Villaseñor L, Ponce H, Brieva J, Moya-Albor E, Núñez-Martínez J, Peñafort-Asturiano C. UP-Fall Detection Dataset: A Multimodal Approach. Sensors. 2019; 19(9):1988. https://doi.org/10.3390/s19091988

[7] Martínez-Villaseñor, L., Ponce, H., Brieva, J., Moya-Albor, E., Núñez-Martínez, J., & Peñafort-Asturiano, C. (2019). UP-Fall Detection Dataset: A Multimodal Approach. Sensors (Basel, Switzerland), 19(9), 1988. https://doi.org/10.3390/s19091988

[8] Ekram Alam, Abu Sufian, Paramartha Dutta, and Marco Leo. 2022. Vision-based human fall detection systems using deep learning: A review. Comput. Biol. Med. 146, C (Jul 2022). https://doi.org/10.1016/j.compbiomed.2022.105626

[9] Shobana, J., Murali, M. An efficient sentiment analysis methodology based on long short-term memory networks. Complex Intell. Syst. 7, 2485 – 2501 (2021). https://doi.org/10.1007/s40747-021-00436-4

[10] Ludwig, Katja & Kienzle, Daniel & Lienhart, Rainer. (2022). Recognition of Freely Selected Keypoints on Human Limbs.

[11] Zheng, Ce & Wu, Wenhan & Yang, Taojiannan & Zhu, Sijie & Chen, Chen & Liu, Ruixu & Shen, Ju & Kehtarnavaz, Nasser & Shah, Mubarak. (2020). Deep Learning-Based Human Pose Estimation: A Survey.

# Fall Classification Using IMU sensor based on Machine Learning Methods

*Sokea TENG[1], Jung-yeon KIM [2], Yunyoung NAM[3]*

*[1]Department of ICT Convergence, Soonchunhyang University, Asan, 31538, Republic of Korea*

*[2] ICT Convergence Research Centre, Soonchunhyang University, Asan, 31538, Republic of Korea*

*[3] Department of Computer Science and Engineering, Soonchunhyang University, Asan, 31538, Republic of Korea*

**Contact: ynam@sch.ac.kr*

*Abstract*— **Fall-Sense: is a cutting-edge project that employs machine learning techniques along with wearable sensors to automatically detect and classify falling types. Falls can happen to anyone and are a crucial area of research for reliable detection and understanding of fall directions. This study we introduce a robust system using 12 wearable sensors on body such as head, shoulder left, shoulder right, upper arm left, upper arm right, forearm left, forearm right, Pelvis, Right Upper Leg, Left Upper Leg, Right Lower Leg, and Left Lower Leg. By using built-in accelerometers and gyroscopes, it accurately captures fall data, distinguishing between 'Non-fall' and 'Fall' events. Furthermore, it classifies the direction of falls as 'Forward-fall,' 'Backward-fall,' or 'Lateral-fall.' In this study we proposed state-of-the-art supervised machine learning method, specifically such as Random Forest (RF), Logistic Regression (LR), Support Vector Machin (SVM), Naïve Bayes, Decision Tree (DT), and k Nearest Neighbors (KNN), we've created a proficient model to classify falls. Our dataset, collected from fall experiments, ensures accurate training. This innovative approach holds the promise of improving safety in various environments. Falls and fall direction classification, enables quick responses, reducing the risk of injuries, particularly for seniors and patients who need constant monitoring and help classify fall type to other researcher want to study on fall risk assessment with long term data.**

## I. INTRODUCTION

Falls are defined as sudden and unintentional collapses from an upright position when a person's legs can no longer support them. These incidents can result in significant physical and emotional harm, disability, and a loss of independence. Additionally, falls may lead to post-fall syndrome, characterized by dependence, loss of autonomy, depression, and further limitations in daily activities, sometimes even resulting in premature death. It's worth noting that falls are not exclusive to the elderly or unhealthy individuals; they can happen unexpectedly to anyone, anywhere.

Recent reports from the World Health Organization (WHO) highlight falls as the second leading cause of unintentional injury deaths worldwide. Each year, approximately 684,000 people globally lose their lives due to falls, with over 80% occurring in low and middle-income countries. Fatal falls are most common among adults over 60 years of age. Furthermore, there are an estimated 37.3 million falls each year that require medical attention [1].

In response to these alarming statistics, wearable sensor devices such as accelerometers, gyroscopes, and pressure sensors have emerged. These devices can capture gait-related data and extract features that help predict fall risks and prevent falls [2]. Wearable sensors automatically record and analyse falling events.

Human Activity Recognition seeks to classify muscle activities and capture physiological data in a timely manner through pervasive computing. This not only contributes to medical diagnosis but also advances research in human activity [3]. As the social issue of aging continues to grow, interest in the activities of daily living (ADLs) among the elderly and related healthcare research is rapidly increasing.

IMU (Inertial Measurement Unit) sensors play a crucial role in collecting data from accelerometers, gyroscopes, and magnetic fields. These sensors are strategically placed in 12 locations, including the head, shoulders, upper arms, forearms, pelvis, and legs. In this study, we propose Random Forest and Naive Bayes models based on automatic machine learning techniques to classify the direction of fall events with high accuracy. This system not only provides insights into when falls occur but also offers valuable information for physiotherapists and clinicians, enhancing the quality of treatments and care. Other primary objectives are to optimize sensor reliability, guarantee user comfort, and ensure that the system is suitable for both patients and healthy individuals while maintaining the highest levels of accuracy.

## II. RELATED WORK

Numerous prior investigations have delved into the classification of falls using wearable sensors and machine learning techniques. For instance, Son et al. (2022) utilized

inertial measurement unit (IMU) sensors in their study, targeting the classification of falls and activities of daily living (ADLs) among agricultural workers. They devised a method to discern fall and non-fall movements by extracting features such as the Euclidean norm of acceleration in both horizontal and vertical planes. Additionally, they employed statistically based methods on nine variables, including mean, standard deviation, variance, maximum, minimum, range, kurtosis, skewness, and correlation coefficient. To distinguish falls from ADLs, they leveraged k-nearest neighbors (kNN) and support vector machine (SVM) algorithms, achieving remarkable ROC AUC-scores of 0.999 in binary-class classification [4].

In the early stages of ADL research, data collection predominantly relied on questionnaires, presenting challenges in terms of sample size and analysis. To overcome these limitations, a user-friendly automated system was introduced for collecting ADL data from elderly individuals during their daily routines. This system incorporated a simplified ADL classification sensor network. Feature extraction was accomplished through a Bayesian network model, while classification employed a Hidden Markov Model (HMM) [5]. Notably, this study extended its scope to include falls as one of the ADL movements, conducting multiple experiments to differentiate falls from various other ADL activities.

In 2013, Nam et al. harnessed a Triaxial Accelerometer and a Barometric Pressure Sensor in conjunction with machine learning techniques for classification purposes. Their study focused on collecting data from children aged 16 to 29 months, with the objective of recognizing daily activities and extracting features, such as mean, standard deviation, and slope of time-domain features calculated over sliding windows. Impressively, their approach achieved an overall activity recognition accuracy of 98.43%. This high accuracy was attained using only a single wearable triaxial accelerometer sensor and a barometric pressure sensor, coupled with a support vector machine [6].

Jennifer Howcroft and her colleagues proposed an innovative approach to enhance fall classification by leveraging sensor data, specifically pressure-sensing insole gait data and a wearable accelerometer. Their research involved a sample of 100 older adults, all aged 75 or older. To refine their analysis, they applied various feature selection algorithms, including Relief-F, Fast Correlation Filter (FCBF), and Correlation-based Feature Selection (CFS). For classification, they employed Naïve Bayesian, Support Vector Machine (SVM), and Multi-layer Perceptron Neural Network classifiers, implementing a 75:25 single stratified holdout approach with repeated random sampling. Their evaluation yielded an accuracy of 78%, sensitivity of 26%, specificity of 95%, F1-score of 0.36, and MCC (Matthews Correlation Coefficient) of 0.31. These results demonstrated that feature selection not only improved classification performance but also reduced the feature sets. To validate these findings, further testing with new data was planned, particularly focusing on the older population [7].

## III. METHODOLOGY

Recently, there're many papers were studied and developed with various methods for fall risk assessment using wearable sensors. The range of supervised machine learning commonly utilized to classifies and analysis sensor data that collected from wearable sensor based on fall risk such as Logistic Regression, Decision tree, Support Vector Machines, and Neural Networks. These techniques can be trained on labelled data to learn the patterns associated with falls and non-falls. In addition, another study was conducted on unsupervised learning algorithm to identify pattern of falls risk.

### *A. Material and Data collection*

This section deals with materials and methods applied in this study. The project was conducted on the application of a classification protocol with a trained data from an IMU (Inertial Measurement Unit) sensor using machine learning techniques to process the obtain dataset in order to classify what kind of falls event.

To record the data of falls, there are three sensors from IMU such as Accelerometer, Gyroscope, and Magnetic Field that each of these consist of three axes (X, Y, Z) for movement measuring records. The data were collected from participants as they completed the performance.

The above sensors were attached on 12 wearables part on the body such as head, shoulder left, shoulder right, upper arm left, upper arm right, forearm left, forearm right, Pelvis, Right Upper Leg, Left Upper Leg, Right Lower Leg, and Left Lower Leg. The obtained dataset we divided into four classes which is each of class represented Non-Fall, Backward Fall, Forward Fall, and Lateral (side). The detail shown on Table 1:

| Falls Type | Files each class | Maximum of row each file | Total samples of each class |
|---|---|---|---|
| None fall | **195** | 600 | 117,000 |
| Backward | **188** | 600 | 112,800 |
| Forward | **202** | 600 | 121,200 |
| Side | **159** | 600 | 95,400 |
| Total | **744** | 600 | 446,400 |

*Table 1: show the numbers of file in each class, numbers of maximum row of each file, and total sample of each class.*

### *B. Data Pre-processing*

In fact, there will be noisily occurred while processing the data collection that may interferes to the model performance. An additional method on pre-processing including noise reduction that using 1D Gaussian filter were proposed in this study to solves this problem. And other we labelling data for classification perform. In dataset, all of files are maximum has 600 rows (600 samples per file).

1). *Noise Reduction*:

The dataset was obtained in time series signal that consists of lots of noise caused by unnecessary movement. The effect of the jittering noise can be reduced by 1D Gaussian filter. The 1D Gaussian filter that operated for smoothly data each of columns. The Gaussian distribution in 1-D's equation as:

$$G(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}} \quad (1)$$

where σ is the standard deviation of the distribution.

applied sigma 1 on all the data that it reached smoothly like that. Figure 1 below is showed the data has applied gaussian filter method:

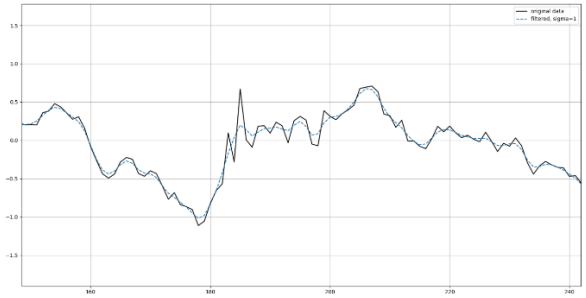


*Figure 1: the graph that showed about data was applied gaussian filter 1D*

2). *Data Labelling*:

According to our dataset has maximum 600 rows of each file, that it is like window size that one sequence has 600 samples. The data was classified and labelled into four groups which deals with supervised machine learning. Each of group are "0", "1", "2", "3" represented "Non-fall", "Backward fall", "Lateral (side)", and "Forward fall", respectively.

3). *Features Extraction:* Our falls classification model, developed through extensive emulation experiments, was constructed using a dataset structured around IMU sensor data. This dataset encompasses readings from three key sensors within the IMU: the 3-axis Accelerometer, 3-axis Gyroscope, and 3-axis Magnetic Field sensors. The dataset creation process involved employing a novel methodology [8][9].

We defined $L2_{norm}$ (or $S_{norm}$) as the L2 norm (also known as the Euclidean norm) of acceleration in the 3-axes and we can calculate via $S_{norm} = \sqrt{{s_x}^2 + {s_y}^2 + {s_z}^2}$, it used for explain the spatial variation of 3 axes acceleration during body movement. Additionally, we defined $S_{hori}$ as a measure of combined acceleration in the horizontal plane (X, Z-plane), determined by taking the Euclidean norm of the acceleration components: $S_{hori} = \sqrt{{s_x}^2 + {s_z}^2}$. Shori quantifies spatial variations in acceleration during movements within the body's horizontal plane. Conversely, $S_{verti}$ represents the Euclidean norm of acceleration in the vertical plane (Y, Z-axis) and is computed as the square root of the sum of squares: $S_{verti} = \sqrt{{s_y}^2 + {s_z}^2}$. Sverti characterizes spatial variations in acceleration during movements within the body's vertical plane. Altogether, these calculations resulted in nine distinct variables, to which we added supplementary statistical features including mean, standard deviation (std), interquartile range (iqr), kurtosis (Kur), skewness (Ske), and correlation coefficients ( $\rho$) [13]. Additionally, we proposed incorporating Energy features, which are commonly utilized in machine learning for signal processing tasks, notably in domains like speech recognition. In our study, we applied Short-Time Energy (STE) [14] and Root Mean Square Energy (RMSE) [15] calculations based on these nine variables. Consequently, each sensor channel contributed to a feature vector comprising 40 independent variables. These variables collectively represent the total number of samples in our dataset, as summarized in Table 2.

| Features | Features Description |
|---|---|
| Mean (x,y,z) | Mean for x, y and z axis |
| Mean (norm, hori, verti) | Mean for norm, hori and verti axis |
| Standard Deviation (x,y,z) | Standard Deviation for x, y, and z axis |
| Standard Deviation (norm, hori, verti) | Standard Deviation for norm, hori and verti axis |
| Kurtosis (x, y, z) | Kurtosis for x, y, and z axis |
| Kurtosis (norm, hori, and verti) | Kurtosis for norm, hori and verti axis |
| Skewness (x, y, z) | Skewness for x, y, and z axis |
| Correlation Coefficient of x,y | Correlation Coefficient between x and y axis |
| Correlation Coefficient of x, z | Correlation Coefficient between x and z axis |
| Correlation Coefficient of y, z | Correlation Coefficient between y and z axis |
| Correlation Coefficient of norm, verti | Correlation Coefficient between norm and verti axis |
| Short-Time Energy (x, y, z) | Short-Time Energy for x, y, and z axis |
| Short-Time Energy (norm, hori, verti) | Short-Time Energy for norm, hori, and verti axis |

*Table 2: table description of features extracted for Machine Learning.*

After extracted we was perform finding the best feature to optimize out system. Table 3 will show you the comparison using our machine learning approached.

| Feature Name | Test Accuracy (%) with 3 sensors | | | | | |
|---|---|---|---|---|---|---|
| | RF | LR | DT | SVM | KNN | NB |
| Mean | 69.20 | 74.55 | 50.45 | 78.13 | 66.07 | 57.59 |
| Std | 68.75 | 74.11 | 49.55 | 75.89 | 67.41 | 58.93 |
| Kur | 60.71 | 61.16 | 45.98 | 58.04 | 53.57 | 50.45 |
| Ske | 58.48 | 63.83 | 43.30 | 64.29 | 60.71 | 56.25 |
| iqr | 63.39 | 62.50 | 40.62 | 64.73 | 51.34 | 56.69 |
| $\rho$ | 55.35 | 54.91 | 42.41 | 61.16 | 48.21 | 51.34 |
| STE | 25.44 | 25.44 | 25.44 | 27.23 | 29.91 | 21.43 |
| RMSE | 25.45 | 25.45 | 25.45 | 25.45 | 25.55 | 23.66 |
| Mean, Std, Kur, Ske, iqr, $\rho$, STE, RMSE | 84.82 | 91.07 | 53.13 | 92.41 | 80.35 | 81.70 |

*Table 3: features and Machine Learning method with Norm, Horizontal, and Vertical vector data.*

From this process, we extracted nine primary variables, complemented by additional statistical features, including mean, standard deviation, STE, RMSE, Kurtosis measures, Skewness measures, Interquartile Range measures, and Correlation coefficients and we have additional statistic method that transpose to iterate over columns as Mean, Standard Deviation, and interquartile range to improve accuracy of our performed, all derived from these nine primary variables. Consequently, the feature vector extracted from a single sensor channel consisted of independent variables. So, the features that prepared for perform with data frame are 216, 324, 432, 540, and 1080 respectively. The dataset was separated such as that detail on the table 4:

| **Name dataset** | **Column of each file** | **Features extraction** |
|---|---|---|
| NIA_2_Wearable (Pelvis and Right Lower Leg) | 18 | 216 |

| | | |
|---|---|---|
| NIA_3_Wearable (Pelvis, Right Lower Leg, and Left Shoulder) | 27 | 324 |
| NIA_4_Wearable (Pelvis, Right Lower Leg, Left Shoulder, and Right Shoulder) | 36 | 432 |
| NIA_5_Wearable (Pelvis, Right Lower Leg, Left Shoulder, Right Shoulder, and Left Forearm) | 45 | 540 |
| NIA_12_Wearable | 108 | 1080 |

*Table 4: show the numbers of columns in each file and numbers of features extraction of each data frame.*

### C. Data Splitting

After preprocessing the data, it is partitioned into two distinct datasets: the training dataset and the testing dataset. In this study, the training dataset is designated to comprise 70% of the entire dataset, leaving the remaining 30% for testing purposes. The subsequent code snippet demonstrates the utilization of the 'train_test_split' function to randomly divide the data into these two subsets.

### D. Data Scaler

As datasets commonly exhibit variations, the necessity for data scaling arises to ensure consistent evaluation. Given the substantial variability in raw data values, effective functioning of objective functions within certain machine learning algorithms relies on normalization. After successfully scaling the data, it undergoes fitting and transformation, as illustrated below.

### E. Machine Learning Algorithms

Machine learning algorithms play a crucial role in recognizing patterns and relationships within sensor data, facilitating the classification of fall risk levels. Among the widely employed machine learning techniques for assessing fall risk, the following stand out:

1). Random Forests (RF): The Random Forest algorithm belongs to the realm of ensemble learning, leveraging multiple decision trees to formulate predictions. In this process, multiple decision trees are trained, with each tree learning from a random subset of the dataset. The model's ultimate prediction is obtained through the aggregation of predictions from all the constituent decision trees.

2). Logistic Regression (LR): Within the spectrum of supervised machine learning approaches applied to sensor data analysis and fall risk identification, methods encompassing decision trees, support vector machines, and neural networks have been employed. Logistic regression, in particular, models data by fitting a logistic function and estimating the probability of data points belonging to respective classes.

3). K-Nearest Neighbors (KNN): As a straightforward yet highly effective classification algorithm, KNN operates by identifying the K closest data points to a given input and assigning it to the majority class within this neighborhood.

4). Support Vector Machines (SVM): In the realm of fall risk assessment, a variety of supervised machine learning methods have been applied to scrutinize sensor data, encompassing decision trees, support vector machines, and neural networks. SVM, for instance, endeavours to identify a hyperplane that optimally segregates data into two classes, maximizing the margin between them [10].

5). Decision Tree (DT): This algorithm constructs hierarchical tree branches, where each branch effectively represents an if-else statement. Branches are created through dataset partitioning based on the most significant features. Ultimately, the final classification occurs at the leaves of the decision tree [11].

6). Naive Bayes (NB): This algorithm relies on Bayes' Theorem, offering a method to compute conditional probability using prior knowledge and the simplistic assumption that each feature is independent of the others. One of its notable strengths is its ability to perform reasonably well even with limited training data, a trait that sets it apart from many other machine learning algorithms [12].

In this study, we conducted an evaluation using several machines learning algorithms, including Random Forest (RF), Logistic Regression (LR), Naive Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), and k-Nearest Neighbors (KNN), to classify falls event.

The Random Forest algorithm falls under the category of ensemble learning, where it leverages multiple decision trees to make predictions. In this process, we train numerous decision trees, each learning from a random subset of the dataset. The model's final prediction is obtained by aggregating the predictions made by all these individual decision trees. A notable practice in Random Forest is to vary the number of estimators (trees) within a specified range, typically from 100 to 800. This variation helps us assess how the number of trees impacts the model's ability to capture feature dependencies and improve classification accuracy.

## IV. RESULTS AND DISCUSSION

### A. Distributions of Data Sample

We need to assess the balance and imbalance within each class. The class counts range from 159 to 202, and according to our calculations using Shannon Entropy, it is approximately 1.997. Therefore, we anticipate that this dataset should exhibit a balanced distribution among the four classes. Additionally, to realistically simulate a fall, it is essential for the subjects to first engage in walking before experiencing the fall, which replicates natural fall scenarios. Consequently, this results in a shortage of data for lateral (side) conditions compared to falls. A visual representation of this data balance is presented in Figure 2.

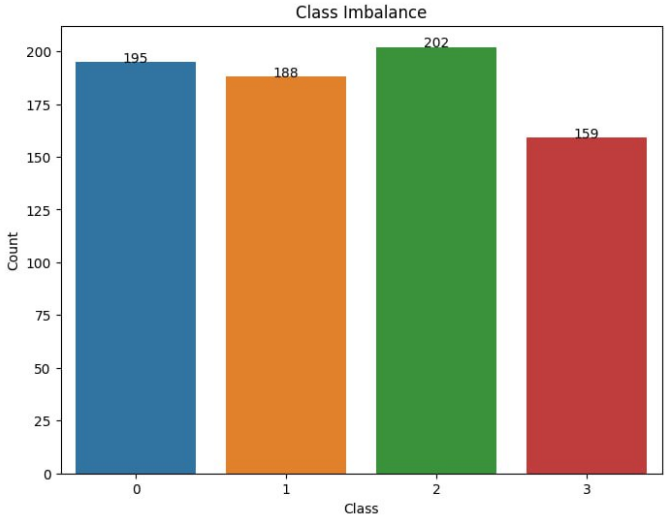



*Figure 2: Imbalance data distributions between "0. normal", "1. Backward-fall", "3. Forward-fall", and "3. Lateral (side)"*

*B. Performance Evaluation*

In this study, we conducted a comprehensive evaluation of our machine learning model using a variety of classification algorithms, including Random Forest (RF), Logistic Regression (LR), Naive Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), and k-Nearest Neighbors (KNN) methods. Our evaluation criteria were based on metrics (2), (3), (4), and (5). which include accuracy, recall, precision, and F_score.

The results from testing our model on a partition of the dataset revealed exceptionally high accuracy. This high accuracy can be attributed to the distinct patterns of tilt angle variation observed in each class, distinguishing between various normal, backward-fall, forward-fall, and lateral (side) events. As a result, these distinct features capturing safe and unsafe behaviours contribute significantly to the overall effectiveness of our falls classification system.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$F_{score} = \frac{Recall \times Precision}{Recall + Precision} \times 2 \quad (5)$$

In the formulas used for this evaluation, 'TP' represents True Positives, 'FN' denotes False Negatives, 'FP' stands for False Positives, and 'TN' signifies True Negatives.

*C. Results*

Based on the information presented in Table 3, it is evident that certain features exhibit high accuracy and are well-suited for the proposed models. These features include mean, standard deviation, interquartile range, kurtosis measures, skewness measures, correlation coefficients, Short-Time Energy (STE), and Root Mean Square Energy (RMSE). Consequently, for our study, we have selected these specific features to be employed in conjunction with our dataset.

In our primary objectives are to optimize sensor reliability, guarantee user comfort, and ensure that the system is suitable for both patients and healthy individuals while maintaining the highest levels of accuracy. So, we have separated dataset to 5 objects that was show in table 3. Now we will let prepare the accuracy of each dataset was separated with accuracy of each wearable places that shown in the table 5.

| Name Models | Test Accuracy (%) | | | | |
|---|---|---|---|---|---|
| | 2 Wearable | 3 Wearable | 4 Wearable | 5 Wearable | 12 Wearable |
| RF | 81.25 | 83.48 | 83.48 | 81.25 | 86.16 |
| LR | 82.14 | 86.16 | 87.95 | 85.27 | 91.07 |
| DT | 57.14 | 60.71 | 63.39 | 58.48 | 55.36 |
| SVM | 85.71 | 87.05 | 87.05 | 90.63 | 92.41 |
| KNN | 75.45 | 80.80 | 77.68 | 76.34 | 80.36 |
| NB | 74.55 | 73.66 | 75.45 | 75.00 | 81.70 |

*Table 5: show the accuracy result that compared wearable sensor place that wear on body.*

In Table 5, we present the results of our evaluation across five distinct objectives, each representing the placement of wearable sensors on the body.

2_Wearable Sensors: This objective focuses on sensors placed on the pelvis and the right lower leg. Notably, SVM achieved an accuracy of 85.71%, RF reached 81.25%, and LR scored 82.14%, demonstrating robust performance.

3_Wearable Sensors: Here, we extended the sensor placement to include the pelvis, right lower leg, and left shoulder. SVM exhibited the highest accuracy at 87.05%, outperforming the 2_Wearable Sensors objective by a margin of 1.34%.

4_Wearable Sensors: This configuration encompasses wearables on the pelvis, right lower leg, left shoulder, and right shoulder. Remarkably, LR yielded the highest accuracy of 87.95%. In a comparison of SVM accuracy, 4_Wearable Sensors came close to the performance of 3_Wearable Sensors, indicating that one sensor could potentially be eliminated without a significant loss in accuracy.

5_Wearable Sensors: With wearables placed on the pelvis, right lower leg, left shoulder, right shoulder, and left forearm, this objective achieved a high accuracy of 90.63% using SVM, surpassing 3_Wearable Sensors by approximately 4.92%.

12_Wearable Sensors: This comprehensive configuration involves wearables positioned on the head, left shoulder, right shoulder, left upper arm, right upper arm, left forearm, right forearm, pelvis, right upper leg, left upper leg, right lower leg, and left lower leg. Impressively, SVM reached an outstanding accuracy of 92.41.07%, surpassing the accuracy of all other objectives. This underscores the remarkable accuracy achievable by utilizing a full set of 12 wearables.

| | | RF | LR | DT | SVM | KNN | NB |
|---|---|---|---|---|---|---|---|

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 2 Wearables | N | 93 | 91 | 77 | 96 | 95 | 84 |
| | BY | 87 | 80 | 61 | 87 | 64 | 72 |
| | FY | 80 | 86 | 47 | 86 | 80 | 77 |
| | SY | 60 | 65 | 40 | 68 | 57 | 60 |
| 3 Wearables | N | 91 | 95 | 72 | 96 | 98 | 86 |
| | BY | 85 | 84 | 67 | 85 | 77 | 66 |
| | FY | 86 | 91 | 56 | 92 | 88 | 76 |
| | SY | 65 | 70 | 42 | 68 | 50 | 65 |
| 4 Wearables | N | 91 | 98 | 70 | 93 | 95 | 89 |
| | BY | 87 | 87 | 66 | 86 | 80 | 70 |
| | FY | 82 | 91 | 62 | 91 | 77 | 77 |
| | SY | 70 | 70 | 53 | 70 | 50 | 60 |
| 5 Wearables | N | 96 | 96 | 79 | 95 | 98 | 88 |
| | BY | 87 | 84 | 59 | 97 | 80 | 72 |
| | FY | 82 | 91 | 52 | 91 | 71 | 76 |
| | SY | 50 | 64 | 47 | 75 | 47 | 60 |
| 12 Wearables | N | 98 | 98 | 75 | 96 | 100 | 89 |
| | BY | 92 | 95 | 59 | 97 | 90 | 87 |
| | FY | 83 | 91 | 45 | 91 | 85 | 91 |
| | SY | 65 | 75 | 38 | 82 | 30 | 47 |
| | | RF | LR | DT | SVM | KNN | NB |

*Table **6**: show the sensitive result of all wearable sensor place that wear on body with falls events.*

Based on table 6, we can make some observations and discuss. For algorithm, performance varies across the dataset and metrics, and RF with SVM is consistently performed well across multiple datasets and metrics. For wearable place, we take from 3 place on body, example based on our performance we can choose 3_wearable place, 5_wearable place, and 12_wearable place.

## V. Conclusion

In summary, this paper has presented an effective fall classification method utilizing data from accelerometers, gyroscopes, and magnetic field sensors. We conducted feature extraction in both the time and frequency domains. To enhance the performance of fall classification, we considered a comprehensive set of 12 features. These features encompassed essential statistical measures such as mean, standard deviation, Short-Time Energy (STE), Root Mean Square Energy (RMSE), Kurtosis, Skewness, Interquartile Range, and various correlation coefficients.

Our study compared multiple feature sets to identify the optimal classification method and evaluated their performance across various body placements. The results demonstrated that the SVM, LR, and RF algorithms achieved average overall accuracies of 92.41%, 91.07%, and 86.16%, respectively. Importantly, these accuracies were achieved with reasonable computational complexity, making them suitable for practical implementation.

This research contributes to the advancement of fall event classification, providing valuable insights that can benefit researchers and practitioners alike. It paves the way for long-term data recording and real-time fall detection systems. Additionally, our investigation explored wearable sensor placements to ensure sensor reliability, user comfort, and suitability for both patients and healthy individuals, all while maintaining a high level of accuracy.

## Acknowledgment

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ICAN (ICT Challenge and Advanced Network of HRD) program (IITP-2023-2020-0-01832) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation)

## References

[1] World Health Organization, *Strategies for preventing and managing falls across the life-course*. Accessed:27-April-2021, Available [online].

[2] Velusamy A, Akilandeswari J, Prabhu R. (2023) "*A Comprehensive Review on Machine Learning Models for Real Time Fall Prediction using Wearable Sensor-based Gait Analysis*". Doi:10.1109/ICIRCA57980.2023.10220663

[3] S. Katz, A. B. Ford, R. W. Moskowitz, B. A. Jackson, and M. W. Jaffe, "*Studies of illness in the aged: The index of ADL: A standardized measure of biological and psychosocial function,*" JAMA, vol. 185, pp. 914-919, Sep. 1963.

[4] H. Son et al., "A Machine Learning Approach for the Classification of Falls and Activities of Daily Living in Agricultural Workers," in IEEE Access, vol. 10, pp. 77418-77431, 2022, doi: 10.1109/ACCESS.2022.3190618.

[5] Korbinian Frank, Maria Josefa Vera Nadales, Patrick Robertson, and Tom Pfeifer. 2010. Bayesian recognition of motion related activities with inertial sensors. In Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing - Adjunct (UbiComp '10 Adjunct). Association for Computing Machinery, New York, NY, USA, 445–446. https://doi.org/10.1145/1864431.1864480

[6] Y. Nam and J. W. Park, "Child Activity Recognition Based on Cooperative Fusion Model of a Triaxial Accelerometer and a Barometric Pressure Sensor," in IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 2, pp. 420-426, March 2013, doi: 10.1109/JBHI.2012.2235075.

[7] Howcroft, J., Kofman, J. & Lemaire, E.D. Feature selection for elderly faller classification based on wearable sensors. J NeuroEngineering Rehabil 14, 47 (2017). https://doi.org/10.1186/s12984-017-0255-9

[8] Althobaiti, T.; Katsigiannis, S.; Ramzan, N. Triaxial Accelerometer-Based Falls and Activities of Daily Life Detection Using Machine Learning. Sensors 2020, 20, 3777. https://doi.org/10.3390/s20133777.

[9] Cleland, I.; Kikhia, B.; Nugent, C.; Boytsov, A.; Hallberg, J.; Synnes, K.; McClean, S.; Finlay, D. Optimal Placement of Accelerometers for the Detection of Everyday Activities. Sensors 2013, 13, 9183-9200. https://doi.org/10.3390/s130709183

[10] Akilandeswari, J. and Jothi, G, Performance Comparison of Machine Learning Algorithms that Predicts Students' Employability (November 15, 2017). Proceedings of the International Conference on Intelligent Computing Systems (ICICS 2017 – Dec 15th - 16th 2017) organized by Sona College of Technology, Salem, Tamilnadu, India, Available at SSRN: https://ssrn.com/abstract=3134357 or http://dx.doi.org/10.2139/ssrn.3134357

[11] R. Kohavi. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. in Proc. 14th Int. Joint Conf. Artif. Intell., Francisco, CA, USA: Morgan Kaufmann Publishers Inc., [Online]. vol. 2, pp. 1137–1143 or Kohavi, Ron. (2001). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. 14.

[12] John, George & Langley, Pat. (2013). Estimating Continuous Distributions in Bayesian Classifiers. Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence.

[13] Cleland, I.; Kikhia, B.; Nugent, C.; Boytsov, A.; Hallberg, J.; Synnes, K.; McClean, S.; Finlay, D. Optimal Placement of Accelerometers for the Detection of Everyday Activities. Sensors 2013, 13, 9183-9200 doi: https://doi.org/10.3390/s130709183

[14] Proakis, J.G., Manolakis, D.G. (2013) Book "Digital Signal Processing" doi: https://books.google.co.kr/books?id=ZoISngEACAAJ

[15] Robert S., and Witte, John S. (2017) Book "Statistics". doi: https://books.google.co.kr/books?id=KcxjDwAAQBAJ

# Enhancing Fall Classifications in Challenging Environments Using Transfer Learning-based Feature Extractions and Classifications

Nab Mat[1], Awais khan [2] Chomyong Kim[3], Seungmin Rho[4], and Yunyoung Nam[5]
[1] *Department of ICT Convergence, Soonchunhyang University, Asan, South Korea*
[2] *Department of ICT Convergence, Soonchunhyang University, Asan, South Korea*
[3] *ICT Convergence Research Centre, Soonchunhyang University, Asan, South Korea*
[4]*Department of Industrial Security, Chung-Ang University, Seoul 06974, South Korea*
[5] *Department of Computer Science and Engineering, Soonchunhyang University, Asan, South Korea*

*Abstract*— **The highly reliable model inference of fall detection in complex environments introduces significant challenges, especially for vulnerable individuals, such as the elderly living alone at home. This paper introduces an innovative approach that harnesses video-based features, utilizing a carefully designed machine learning mode, to enhance fall detection using a single camera view. We proposed a novel approach that leverages transfer learning-based feature extraction combined with a machine learning classification model to enhance the accuracy and reliability of fall detection. Our method starts with the data preparation of the frame with the extraction of discriminative features from the video data frame by using a pre-trained transfer learning model. The extracted feature has a powerful representation of the underlying dynamic of fall events. Subsequently, we employed the machine learning classifier to perform the extracted feature and classifier the fall events. Additionally, we evaluated our approach on a diverse dataset containing various scenarios, demonstrating its effectiveness compared to existing methods. By utilizing transfer learning and machine learning, we aim to enhance the safety and reliability of fall detection events in challenging tasks.**

## I. INTRODUCTION

Falling is the main global issue among older adults over the age of 65 who are living alone at home and people of all age groups due to various causes, including weakening of the body, previous history of falling, use of assistive devices, issues with balance or walking, trouble seeing or hearing poorly lit environments [1]. Therefore, falling detection has become an extensively researched topic with several utilizing computerization methods to analyze them. Hence, contribution finding is needed to use the novelty algorithm for fall detection.

There are various types of falling that cause injuries such as loss of balance, falls when sitting down, backward fall, forward fall, and fall by side. These kinds of fall by direction. In addition, physical activities and cardiovascular disorders cause falls [1-2].

According to the Public Health Agency of Canada, in 2026, one Canadian older than 65 will be out of five whereas in 2001 the portion was eight to one. It is notable that 93% of elderly people stay in their private house and 29% of them will be led to live a lonely life [3]. To detect falling, recently, there have been several proposed methods and they used high-performance devices such as wearable sensors, gyroscopes, accelerometers, magnetic sensors, cameras, and so on. However, there still are limitations due to the amount of data and the utilizing devices' problems for wearable devices such as user compliance issues and discomfort, battery life, installation and maintenance, and limited coverage. Furthermore, the camera's issues are data resolution, scalability, data synchronization in case of using multiple cameras, and input feature limitation. Furthermore, a fall accident may not cause significant health issues, but healthy people are still at risk of accident falling and typically, they do not have access to certain devices that require expensive 24-hour monitoring [4]. Therefore, there is also computer vision proposed for fall detection by using camera video.

Even for the fall detection using the camera in the existing work has been done with the various approaches, but it still faces several limitations. To begin with, cameras have a limited field of view, quality of background's light set up, distances between objects and cameras, missing falls accruing outside their fall coverage area, and complexity of human action which is a challenging issue in the accurate detection of fall due to most of recently researched focus on the fall and non-fall only [5]. Falls among the elderly population have become a growing concern due to their potential to cause severe injuries and a decline in overall health and independence.

We proposed a machine learning model by utilizing a single camera with the transfer deep learning model for fall detection. It is passionate about dealing with computer vision because it does not require the person to wear anything on

their body. We have been collecting our own dataset that has high quality with the high capability of environment setup. Moreover, we aim to conduct four classes that are fall-backward, fall-forward, fall-sideway, and non-fall. We have selected the dataset from two healthy participants of adults in the hospital who performed the video within 10 seconds per video. To enhance the accuracy of our studies, we will utilize data from a single camera viewpoint as an input feature. By incorporating information from all the frame that has been extracted from one camera view, we aim to create more comprehensive and robust input features after using a pre-trained model.

Pre-trained deep learning models refer to the first model that has learned a lot from a large dataset already. Instead of creating the new model from scratch, then from this model, we can use it to train another model on a different task or dataset [6]. This approach led to our system being better at recognizing complex patterns in the data from a single camera.

By combining these techniques, this approach is poised to offer a more nuanced and accurate understanding of the environment, potentially yielding superior results in the domain of fall detection. This study is separated into five sections. In section number II, we present the related work. In section III, we present about the proposed methodology for falling detection by single camera viewpoint. In section IV, we present the experiment results and discussion of the system. We ended this work in section V with the conclusion and future research.

## II. Related Work

In recent years, the falling issue has emerged as an important role area of research. Falls accidents may not cause significant health issues, but healthy people are still at risk of accident falling and typically, they do not have access to certain devices that require expensive 24-hour monitoring. In response to this pressing issue, researchers and engineers have dedicated their efforts to developing innovative fall detection systems based on camera video aimed at promptly identifying and alerting caregivers or medical professionals when a fall occurs. In the realm of fall detection research, numerous studies have contributed to advancing our understanding of this critical area.

Wu et al. [7], the authors present a comprehensive investigation into the design and implementation of a fall detection system utilizing wearable sensors. A wearable device positioned at the individual's waist and the system can detect the elderly's falling by acceleration analysis.

Xiaoqun Yu et al. [8], offer a significant contribution to the domain of pre-impact fall detection. They introduce the "KFall" dataset, a comprehensive resource developed from the activities of 32 Korean participants engaging in 21 types of activities of daily living (ADLs) and simulating 15 types of falls. This dataset includes rich motion data comprising acceleration, angular velocity, and Euler angles, all collected using a nine-axis inertial sensor positioned at the participants' lower back.

Yhdego et al. [9], conducted a feasibility study outlined in their work titled "Toward Real-Time, Robust Wearable Sensor Fall Detection Using Deep Learning Methods." Their research addresses critical aspects of this field, including the evaluation of different sliding-window segmentation techniques in deep-learning-based fall detection. This study fills a notable gap by systematically exploring various segmentation methods, going beyond conventional sliding windows.

Nait Aicha et al. [10], present that the Early detection of high fall risk is vital for preventing falls in older adults. This study investigates whether deep learning methods can automatically derive fall risk assessment features from raw accelerometer data. A dataset of 296 older adults is used to compare the performance of three deep learning model architectures (CNN, LSTM, ConvLSTM) against a baseline model with biomechanical features.

Delgado-Escano at el. [11], introduces a novel approach rooted in deep learning to address both fall detection and person identification simultaneously. The primary objective is to develop a versatile model capable of seamlessly adapting to various datasets without the need for extensive model parameter tuning. the employment of a k-NN classifier efficiently processes these feature vectors for the tasks of fall detection and person identification.

Putra at el. [12], in this study, the author introduces an innovative event-triggered machine learning (EvenT-ML) approach designed to address these limitations. EvenT-ML aligns each fall stage precisely, enhancing the recognition of characteristic features associated with different stages. To assess the effectiveness of this approach, the researchers utilized two publicly accessible datasets and employed a classification and regression tree (CART), k-nearest neighbor (k-NN), logistic regression (LR), and the support vector machine (SVM) to train the classifiers.

## III. Method

In this section, we introduce the methodology outline that we proposed for employing fall detection and data information Fig.1. We have collected the dataset from 2 healthy adults male and female. During the video recording, they must wear the patient clothes and perform the actions of each four classes such as fall forward, fall backward, sideway fall, and non-fall. We have three scenarios of fall which are stumble, loss of balance and fall, and slip and fall. The video's time interval is 10 seconds with 59.94 frames per second.


Fig 2. Sample image in each class

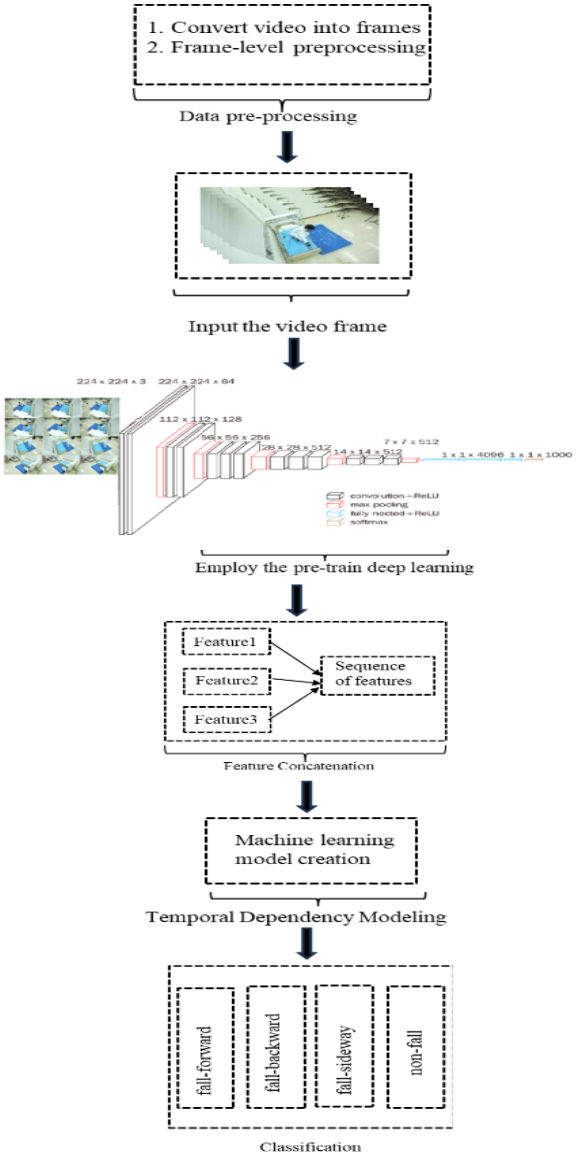



Fig 1. Proposed model structure.

## A. Data pre-processing

In this section, we first work on the data preparation, we convert the video to individual frame, and we downscale the video frame to 600*400 to reduce memory usage. It can make it easier to create visualizations of the data because the original frame is too large. Table 1 shows the dataset information used in this study.

TABLE I
DATASET INFORMATION

| Classes | Total sample | Duration/Frame rate |
|---|---|---|
| 1. fall-forward | 1200 frames | 10s/video<br>60frame/s |
| 2. fall-backward | | |
| 3. fall-sideway | | |
| 4. non-fall | | |
| Totally frame | 4,800 frames | |

## B. Pre-train deep learning

In this stage, we applied the pre-training deep learning model for the feature extraction. A pre-trained deep learning model is a neural network that has been trained on a large dataset for a specific task before being fine-tuned or used for another related task. This approach has become increasingly popular because it can save significant computational resources and time compared to training a deep learning model from scratch. In this study, we applied VGG16 pre-train deep learning model for the fall classification. It is a well-established and widely used architecture for image classification tasks. VGG16 has shown strong performance on various image recognition tasks and has become a popular choice as a feature extractor in transfer learning scenarios.

## C. Feature Concatenation

Feature concatenation is a process in which we combine, or concatenate features extracted from different sources or modalities into a single feature vector. In the context of video frame classification, it typically involves taking the features extracted from individual frames (spatial features) and combining them to create a single feature vector for each video sequence.

## D. Temporal Dependency Modeling

In the domain of fall detection, accurately discerning genuine falls from routine activities is of paramount importance. The 'Temporal Dependency Modelling' step serves as a pivotal component in achieving this objective. It involves the analysis of how events evolve over time within a video sequence. By treating video frames as a temporal sequence, this step recognizes the progression of actions and employs a machine-learning model, to capture temporal dependencies. During training, the model learns to identify patterns and dependencies that signify fall events by considering the evolving context of actions. This temporal perspective significantly enhances the accuracy of fall detection, reducing false alarms, and improving the system's reliability in identifying fall-related incidents. In essence, Temporal Dependency Modelling empowers fall detection systems to comprehend the unfolding narrative of events over time, leading to more precise and dependable fall detection outcomes.

## E. Classification

In the final section of the proposed model, the system leverages the insights gained from Temporal Dependency Modelling to make informed decisions about the nature of observed events within video sequences. Building upon the temporal context established earlier, this step employs machine learning or deep learning models to classify the sequences in each class event. The models are trained to recognize distinct patterns and features indicative of falls, ensuring precise and reliable identification. By using the knowledge acquired through Temporal Dependency Modelling, the 'Classification' step plays a crucial role in enabling fall detection systems to respond to fall-related

incidents accurately and swiftly, ultimately contributing to enhanced safety and well-being in various environments.

## IV. Result

The machine learning models performed the result for the classification of fall events with different outcomes. The performance of our classifiers was assessed using a comprehensive set of evaluation metrics, providing insights into the effectiveness of our models. For the Support Vector Machine (SVM) model performed the result with an accuracy of 92.85%, F1 Score of 93%, recall of 93%, and precision of 94%.

The Random Forest (RF) model demonstrated a high level of accuracy, reaching 97.71% and, the F1-score of 95%. With a recall rate of 95%, the RF model effectively captured positive instances, and its precision of 96% indicated a strong level of accuracy in identifying positive cases.

The k-Nearest Neighbors (KNN) model achieved an accuracy of 82.57%, indicating a level of correctness in its predictions. It also demonstrated an F1 Score of 83%, showing a well-balanced trade-off between precision and recall. The model's recall rate of 83% denoted its ability to effectively identify positive instances, while its precision of 84% indicated a strong accuracy in predicting positive cases.

Lastly, the Decision Trees (DT) model attained an accuracy of 81.39%. It obtained an F1 Score of 79%, recall of 81%, and precision of 84%. These results collectively provide insights into the performance of each classifier and their suitability for classifying fall events.

Below are the equations and descriptions for key performance metrics and tools often used to evaluate and display the results of a classification model:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\_score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

Where: True Positive (TP)

- The predicted value matches the actual value, or the predicted class matches the actual class.
- The actual value was positive, and the model predicted a positive value.

True Negative (TN)

- The predicted value matches the actual value, or the predicted class matches the actual class.
- The actual value was negative, and the model predicted a negative value.

False Positive (FP)

- The predicted value was falsely predicted.
- The actual value was negative, but the model predicted a positive value.

False Negative (FN)

- The predicted value was falsely predicted.
- The actual value was positive, but the model predicted a negative value.

TABLE II
THE COMPARISON OF THE MODEL EVALUATIONS

| Model | Accuracy | F1 Score | Recall | Precision |
|---|---|---|---|---|
| SVM | 92.85% | 93% | 93% | 94% |
| RF | 97.71% | 95% | 95% | 96% |
| K-NN | 82.57% | 83% | 83% | 84% |
| DT | 81.39% | 79% | 81% | 84% |

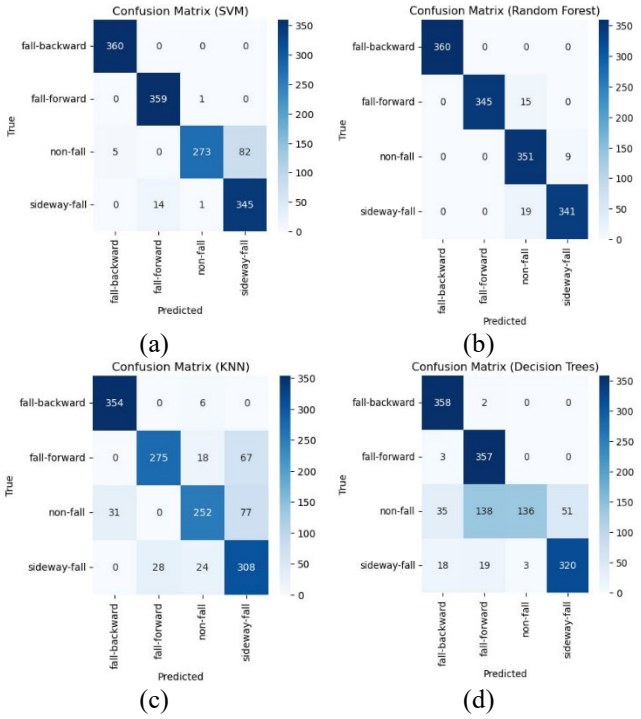



Fig 3. The confusion matrix of each model

## V. Conclusion

In this paper, we have presented a method for falling detection by using a single camera. Our approach pre-training deep learning model processing techniques for data pre-processing and feature extraction from the image frames. In addition, we classify the fall classification by utilizing deep learning algorithms which are long-short-term memory models to train a classification model. Overall, our study is passionate about the fall classification using a deep learning model and we do the classification into four classes which are fall-forward, fall-backward, fall-sideway, and non-fall.

For future work to enhance more accuracy of the fall events prediction, we are willing to increase the sample size of the dataset in each class. In addition, we aim to try multiple pre-trained deep learning models with machine learning models and deep learning models.

## Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2023-2018-0-01799) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)

## REFERENCES

[1] Chandak, Ayush, and Nitin Chaturvedi. "Machine-learning-based human fall detection using contact-and noncontact-based sensors." Computational intelligence and neuroscience 2022 (2022).

[2] Usmani, Sara, Abdul Saboor, Muhammad Haris, Muneeb A. Khan, and Heemin Park. 2021. "Latest Research Trends in Fall Detection and Prevention Using Machine Learning: A Systematic Review" *Sensors* 21, no. 15: 5134.

[3] Sultana, Arifa, Kaushik Deb, Pranab Kumar Dhar, and Takeshi Koshiba. 2021. "Classification of Indoor Human Fall Events Using Deep Learning" Entropy 23

[4] Shu, F., Shu, J. An eight-camera fall detection system using human fall pattern recognition via machine learning by a low-cost android box. Sci Rep 11, 2471 (2021).

[5] Shotkit, Marsupial Drive, Pottsville, NSW Australia

[6] Dipesh Shrestha — Published On September 17, 2022 and Last Modified On October 12th, 2022

[7] Wu, Falin, Hengyang Zhao, Yan Zhao, and Haibo Zhong. "Development of a wearable-sensor-based fall detection system." International journal of telemedicine and applications 2015 (2015): 2-2.

[8] A Large-Scale Open Motion Dataset (KFall) and Benchmark Algorithms for Detecting Pre-impact Fall of the Elderly Using Wearable Inertial Sensors

[9] Yhdego, Haben, Christopher Paolini, and Michel Audette. "Toward Real-Time, Robust Wearable Sensor Fall Detection Using Deep Learning Methods: A Feasibility Study." Applied Sciences 13, no. 8 (2023): 4988.

[10] Nait Aicha, Ahmed, Gwenn Englebienne, Kimberley S. Van Schooten, Mirjam Pijnappels, and Ben Kröse. "Deep learning to predict falls in older adults based on daily-life trunk accelerometry." Sensors 18, no. 5 (2018): 1654.

[11] Delgado-Escano, Ruben, Francisco M. Castro, Julian R. Cozar, Manuel J. Marin-Jimenez, Nicolás Guil, and Eduardo Casilari. "A cross-dataset deep learning-based classifier for people fall detection and identification." Computer methods and programs in biomedicine 184 (2020): 105265.

[12] Putra, I. Putu Edy Suardiyana, James Brusey, Elena Gaura, and Rein Vesilo. "An event-triggered machine learning approach for accelerometer-based fall detection." Sensors 18, no. 1 (2017): 20

# Stability Analysis of Networked Control Systems and Their Applications to Assisted Robots

Seok Young Lee[1,*]
[1]*Electronic Engineering, Soonchunhyang University, Asan, Republics of Korea*
*Contact: suk122@sch.ac.kr

*Abstract*— **This paper is concerned with the stability analysis problem of a network-controlled helicopter system with an asynchronous sampling time. Due to inherent limit of hardware resources, networked control system has sampling times when controlling remote systems. By utilizing an input-delay approach, sampled-data system can be reformulated into a linear time-varying discrete-time system. A novel numerical example which considers a practical helicopter system is discussed.**

## I. Introduction

This paper is concerned with the stability analysis problem of a network-controlled helicopter system via an input-delay approach. Between two sequential sampling times, a sampled-data system is controlled by previously sampled states and thus can be modelled as a time-delay system with a saw-tooth delay. Inevitably, information at the sampling instants have played essential roles in controlling the systems and reducing the conservatism of the stability criteria. This paper utilizes the results of [2] for stability analysis of a network-controlled helicopter system. A novel numerical example for a practical helicopter system demonstrates the effectiveness of [2] in terms of allowable sampling intervals.

## II. Main Results

Consider the following sampled-data system of [1].

$$\dot{x}(t) = Ax(t) + A_d x(t_k), \quad \forall t \in [t_k, t_{k+1}), \qquad (1)$$

where $x(t) \in R^n$ is the system state, $A, Ad \in R^{n\times n}$ are system matrices, and $t_k$ is the sampling instant such that $\cup k \in N[t_k, t_{k+1}) = [0, +\infty)$.

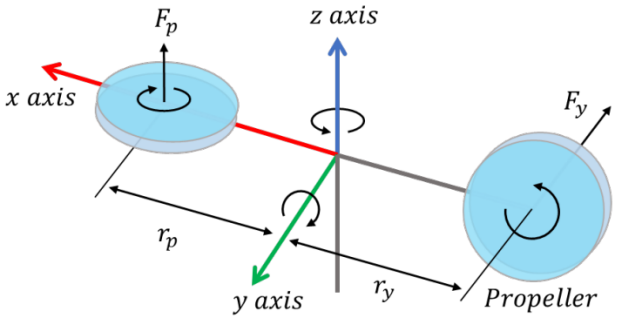



Fig. 1 A free-body diagram of the 2-DOF helicopter system

The sampling interval is defined as

$$t_{k+1} - t_k = h_k \qquad (2)$$

Based on Theorem 2 of the paper [2], this paper newly discusses the Quanser AERO 2-DOF helicopter system [3] with a asynchronous sampling time. This helicopter system can be modelled as a free body diagram in Figure 1. This system consists of two identical rotors that produce the thrust forces $F_p(t)$ and $F_y(t)$ acting at two points with distances $r_p$ and $r_y$ from the z-axis along the x-axis, respectively. Thus, a propeller generates a torque around the y-axis leading to a pitch motion, while the other handles a yaw motion. This system can be described as follows:

$$\tau_p(t) = J_p\ddot{\theta}(t) + D_p\dot{\theta}(t) + K_{sp}\theta(t), \qquad (3)$$

$$\tau_y(t) = J_y\ddot{\phi}(t) + D_y\dot{\phi}(t), \qquad (4)$$

where $J_p = J_y = 0.0215$ are moments of inertia about pitch axis and yaw axis, respectively. $D_p = 0.0071$ and $D_y = 0.0220$ are viscous friction constants about pitch axis and yaw axis, respectively. $K_{sp} = 0.0374$ is the stiffness about the pitch axis. The torques $\tau_p(t)$ and $\tau_y(t)$ which respectively acts on the pitch and the yaw axes are assumed to be proportional to the input DC voltages Vp(t) and Vy(t) of the rotors such that:

$$\tau_p(t) = K_{pp}V_p(t) + K_{py}V_y(t), \qquad (5)$$

$$\tau_y(t) = K_{yp}V_p(t) + K_{yy}V_y(t), \qquad (6)$$

where $K_{pp} = 0.0011$ and $K_{py} = 0.0021$ are thrust torque gain acting on pitch axis from pitch propeller and yaw propeller, respectively. Also, $K_{yp} = -0.0027$ and $K_{yy} = 0.0022$ are thrust-torque gains acting on yaw axis from pitch propeller and yaw propeller, respectively. By utilizing the equations (4)-(6) and the state variable $x(t) = [\theta(t)\ \varphi(t)\ \dot{\theta}(t)\ \dot{\varphi}(t)]T$ , this helicopter system can be represented as a sampled-data system (1) with the following matrices

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -K_{sp}/J_p & 0 & -D_p/J_p & 0 \\ 0 & 0 & 0 & -D_y/J_y \end{bmatrix}, \qquad (7)$$

$$A_d = BK,\ B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ K_{pp}/J_p & K_{py}/J_p \\ K_{yp}/J_y & K_{yy}/J_y \end{bmatrix}, \qquad (8)$$

$$K = \begin{bmatrix} 0.0432 & 0.0530 \\ 1.1617 & -0.6085 \\ -0.1687 & -0.2070 \\ -0.1789 & 0.0937 \end{bmatrix}^T . \quad (9)$$

Here, a matrix $K$ is a gain of a sampled-state feedback controller, and $B$ is a system matrix. In this example, i newly derive an analytic upper bound $h = 12.0942$ of a periodic sampling interval. With a periodic sampling time, sampled-data system can be regarded as a linear discrete-time system. Integrating the differential equation (1) yields

$$x(t) = \Gamma(t - t_k), t \in [t_k, t_{k+1}],$$
$$\Gamma(\tau) = e^{A\tau} + \int_0^\tau e^{A(t-\tau)} dr A\_d\,, \tau \geq 0.$$

Under the periodic sampling $h = h_k$, the system dynamics in (1) becomes

$$x(t_{k+1}) = \Gamma(h)x(t_k) \quad (10)$$

This system is asymptotically stable if and only if $\Gamma(h)$ has all eigenvalues inside the unit circle. However, under asynchronous sampling, such method does not hold. Therefore, utilizing Theorem 2 of the paper [2], i verify that the system (1) with the matrices (7)-(9) is asymptotically stable under $h_k \in [10^{-5}, 9.5042]$. In Figure 2, all state responses with an initial condition $x(0) = [10\ 45\ 0\ 0]T$ converge.

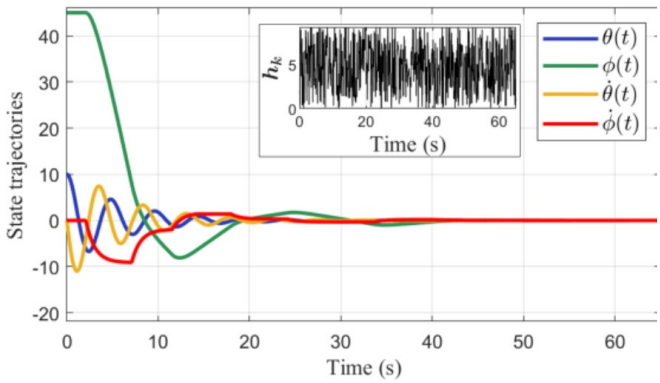



Fig. 2 The state $x(t) = [\theta(t)\ \varphi(t)\ \dot{\theta}(t)\ \dot{\varphi}(t)]T$ trajectories of the helicopter system with an asynchronous sampling interval $h_k \in [10^{-5}, 9.5042]$

## III. Conclusions

This paper has discussed stability analysis of the sampled data controlled helicopter system with a asynchronous sampling time. In the future works, the derived results also can be applied to problems concerned with sampled-data controller synthesis and synchronization of various systems including fuzzy systems, switched systems, and delayed chaotic Lur'e systems.

## References

[1] N. K. Kwon and S. Y. Lee, "Novel equalities for stability analysis of asynchronous sampled-data systems," IEEE Access, vol. 8, pp. 177 195–177 205, 2020.

[2] S. Y. Lee, "Improved stability criteria for sampled data systems via a novel looped-functional," in The 13th Asian Control Conference (ASCC), May 2022.

[3] J.M. Park, "An improved stability criterion for networked control systems with a constant transmission delay," Journal of the Franklin Institute, vol. 359, no. 9, pp. 4346–4365, 2022.

# Development and Verification of Swallowing Disorder Diagnosis Model

DongWook Lim[1], Chung Sub Lee[1], Go-eun Lee[1], Chang Won Jeong [1, 2], Hee Kyung Moon [3,*]
[1]*Medical Convergence Research Center, Wonkwang University, South Korea*
[2] *Center Research Center of Biomedical Research Institute of Wonkwang University Hospital, South Korea*
[3] *Institute for Educational Innovation, Wonkwang University, South Korea*
*Contact: ybnjcw@wku.ac.kr, phone +82-10-5684-0412

*Abstract*—**We suggest an AI diagnosis model for detecting swallowing disorders, including penetration and aspiration. Normal swallowing stages, which can be assessed through imaging, are generally classified into three phases: Oral Phase, Pharyngeal Phase, and Esophageal Phase. Swallowing disorders are further categorized into Penetration and Aspiration, broadly falling into two categories. In this paper, we propose an artificial intelligence (AI) model for diagnosing swallowing disorders. This is achieved by labeling video files of patients with swallowing disorders into these five classes and applying the developed AI model.**

## I. INTRODUCTION

Recently, the method for diagnosing swallowing disorders is the VFFS (Videofluoroscopic Swallowing Study) test. This is one of the important testing methods for diagnosing and evaluating swallowing disorders. This test uses X-rays to observe a patient's swallowing of food or drink in real time to identify any problems. Swallowing can be divided into three stages: the oral stage, the pharyngeal stage, and the esophageal stage. Dysphagia is a result of one or multiple disorders in any of these stages [1,2]. In particular, during the initial stage of swallowing food bolus, there may not be significant issues, but during the pharyngeal stage where the actual swallowing occurs, the epiglottis unconsciously performs the action of blocking the entrance of the airway to prevent the food bolus from entering the trachea and facilitating its natural passage into the esophagus [3]. Dysphagia can be classified into two cases. The first case is aspiration, where food particles do not pass through the vocal folds but linger around the folds and enter the airway, passing through below the vocal folds. The second case is penetration, where food particles do not pass through the vocal folds but enter the airway and go down as far as below the vocal folds.

Therefore, we suggest an AI model to automatically detect the presence of penetration or aspiration in patients with swallowing disorders based on VideoFluoroGraphy (VFG) images.

## II. AI MODEL DEVELOPMENT

In this paper, we propose a swallowing disorder detection model using YOLOv7. For this, we collected the VFS medical images and then generated a training dataset using labeling tools.

### A. Data Collection

Patients with dysphagia according to ICD 10 code were collected from EMR. The participants were between 5 and 96 years old (mean age 68.3 ± 17.8 years) and included 169 males and 80 females. Patients are divided into patients with aspiration findings (N=77), patients with invasive findings (N=141), and patients with difficulty swallowing or moving their body too much (N=31).

### B. Generation of Train Dataset

*Upon examining the total frames of data from 249 patients to date, the technician directly confirmed normal swallowing findings: oral, pharyngeal, and esophageal phases, as well as abnormal swallowing findings: Penetration and Aspiration. Three skilled specialists and technicians independently labeled them. The dataset was created and evaluated progressively. Finally, labeling datasets for AI learning were created for Oral (2355), Pharyngeal (2338), Esophageal (1480), Penetration (1856), and Aspiration (1320).*

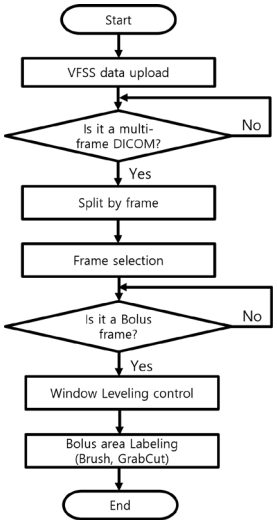



Fig. 1 Swallowing disorder medical image labeling process

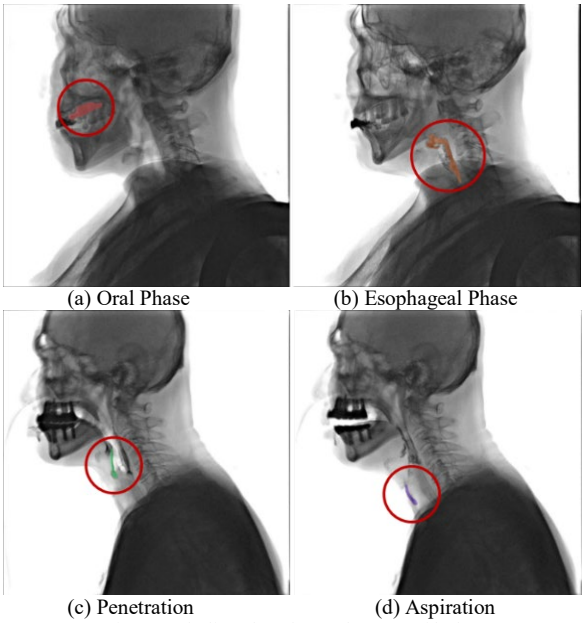
The labeling web application manages VFG images and handles file management. The study images support a multi-frame format to track the entire path of movement from the oral

cavity to the stomach. Upon uploading the file, the Node server saves the original video to storage and converts the multi-frames into images on a per-frame basis for labeling. Subsequently, the application processes and saves the video as an AVI file for swallowing disorder prediction using a detection AI model. If any image processing tasks are requested for labeling in the medical image labeling web application, such as using a Brush tool, they are processed using Python packages in the Flask micro web framework. The entire user interface (UI) of the medical image labeling web application is shown in Figure 2. The labeling tool progresses while being confirmed through a viewer.

Fig. 2 Medical image labeling web application UI

The three stages of swallowing are as follows:

1. The oral stage, which involves chewing food if necessary or creating a consistency and form that can be swallowed in the mouth until the tongue pushes the food back to induce pharyngeal swallowing.
2. The pharyngeal stage, which involves pharyngeal swallowing. This is triggered, and the food lump passes into the pharynx.
3. The esophageal stage, which consists of the phase in which the food lump is moved through the esophagus to the stomach through the peristaltic movement of the esophagus.
4. 



(a) Oral Phase (b) Esophageal Phase

(c) Penetration (d) Aspiration

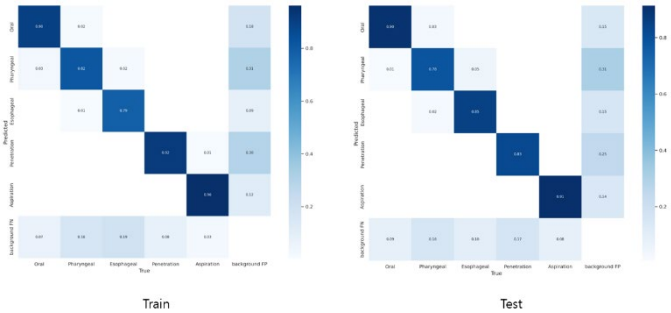
Fig. 3 Labeling data for each stage of bolus

### C. AI Model for swallowing disorder detection

In this study, we developed an artificial intelligence model for Object Detection of 5 Classes (Oral, Pharyngeal, Esophageal, Penetration, Aspiration) using YOLOv7[7]. We utilized four Tesla V100 32GB GPUs on an NVIDIA DGX Station. The total learning time was 14 hours and 10 minutes. In the case of the YOLOv7 model, original-size images were used as learning data, and image data labeled together were used as learning data.

## III. RESULTS

The performance evaluation results of the YOLO model we adopted are depicted in Figure 4. The training data showed accuracy for Oral: 0.90, Pharyngeal: 0.82, Esophageal: 0.79, Penetration: 0.92, Aspiration: 0.96. The test data also exhibited accuracy for Oral: 0.90, Pharyngeal: 0.78, Esophageal: 0.85, Penetration: 0.83, Aspiration: 0.91. The model demonstrated excellent performance in detecting Penetration and Aspiration.



Train Test

Fig. 4 Confusion Matrix for swallowing disorder diagnosis model

## IV. CONCLUSIONS

In this study, we proposed a disease diagnosis model through the automatic detection of various boluses in VFG image data of patients with dysphagia. We demonstrated the model's excellence through verification results. We anticipate that deep learning based on this research can be applied to diagnosing swallowing disorders.

## ACKNOWLEDGMENT

This study was supported by the grant of the National Research Foundation of Korea funded by the Korean government (MSIP) (No. NRF- 2021R1A5A8029876), the grants of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare (HI22C0787).

## REFERENCES

[1] Ariji, Y., Gotoh, M., Fukuda, M. et al. A preliminary deep learning study on automatic segmentation of contrast-enhanced bolus in videofluorography of swallowing. Sci Rep 12, 18754 (2022). https://doi.org/10.1038/s41598-022-21530-8

[2] Stevenson RD, Allaire JH. The development of normal feeding and swallowing. Pediatr Clin North Am. 1991 Dec;38(6):1439-53. doi: 10.1016/s0031 -3955(16)38229-3. PMID: 1945550.

[3] Logemann J. 1983. Anatomy and physiology of normal deglutition. In: J Logemann, editor. Evaluation and treatment of swallowing disorders. San Diego: College Hill Press. p 9–36

# Physical Layer UE Procedure for Autonomous UE Resource Allocation in 5G NR V2X

Daegun. Jang[1], Gayeon. Kim[1], Taejun. Choi[2], and Tahyeong. Kim[*,2]
[1]*ICT convergence, Soonchunhyang University, Asan, Korea*
[2]*Information and communication Engineering, Soonchunhyang University, Asan, Korea*
*Contact: First.wowhensum@sch.ac.kr, phone +82-10 7210 4419. Second.gayeon17@sch.ac.kr, phone +82-10 3646 5917.
Third.20184123@sch.ac.kr, phone +82-10 6309 8087. Fourth.th.kim@sch.ac.kr, phone +82-10 4178 6529

*Abstract*— **Vehicle-to-Everything (V2X) communication technology is a key enabler for realizing future intelligent traffic systems, and it is currently an active area of research in academia and industry. Recently, the 3rd Generation Partnership Project (3GPP) has completed standardization for V2X based on 5G New Radio (NR). In NR V2X, one of the technologies designed to support Vehicle-to-Vehicle (V2V) communication outside the base station coverage area is the Resource Allocation (RA) Mode 2, which allows User Equipment (UE)s to autonomously perform resource allocation. In this paper, we introduce the physical layer standard techniques for the resource allocation procedure of UEs operating in Mode 2 for resource allocation in NR V2X.**

## I. Introduction

Vehicle-to-Everything (V2X) communication technology, which encompasses communication between vehicles and everything else, is a core technology for realizing next-generation intelligent traffic systems, including autonomous driving. Research in various aspects of V2X has been actively conducted in recent times [1], [2]. V2X includes Vehicle-to-Vehicle (V2V) communication between vehicles, Vehicle-to-Pedestrian (V2P) communication between vehicles and pedestrians, Vehicle-to-Infrastructure (V2I) communication between vehicles and nearby infrastructure, and Vehicle-to-Network (V2N) communication between vehicles and networks [3]. With V2X, vehicles can communicate with surrounding vehicles and road infrastructure, enabling real-time sharing of road traffic information and other traffic services. This allows for swift responses to unexpected situations on the road.

Recently, the 3rd Generation Partnership Project (3GPP) completed the first standardization for sidelink (SL), which corresponds to V2V communication for supporting V2X based on 5G New Radio (NR) in Release 16 [4], [5]. 3GPP defined four core V2X use cases as Cluster Driving, Advanced Driving, Remote Driving, and Extended Sensor [6]. To provide each of these services, challenging requirements such as high data rates, low latency, high reliability, and high connection density must be met. To meet these requirements, Release 16 NR V2X has extensively standardized technologies that were not supported in V2X based on LTE (Long Term Evolution), including unicast and groupcast communication, support for high modulation orders, Hybrid Automatic Repeat and Request (HARQ) techniques, Channel State Information (CSI)

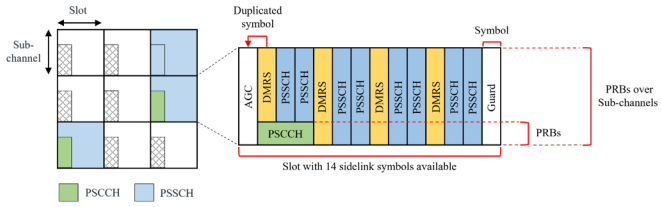



Fig. 1 PSSCH scheduling in NR V2X.

measurement and reporting, among others [7]. NR V2X supports Resource Allocation (RA) Mode 2, where UEs can autonomously select communication resources, even in environments beyond the coverage area of base stations [9]. Mode 2 is essential for ensuring uninterrupted vehicle-to-vehicle connectivity in scenarios where there is no base station or base station control, such as complex urban environments and remote areas. However, in Mode 2, there is a potential issue of signal interference when different UEs select the same resources for transmission [1], [9]. To address this issue, Release 16 NR SL has adopted resource sensing and selection techniques as standardized technologies, allowing UEs to sense interference levels on specific resources and avoid interference when performing transmissions [6], [9].

In this paper, we introduce the physical layer standard specifications for the RA procedure of UEs operating in Mode 2, based on 3GPP Release 16 SL standards. To do this, we first analyze the NR V2X physical layer transmission structure and then delve into the standardized techniques for resource sensing and resource selection procedures. Additionally, we introduce resource re-evaluation procedures to address inaccuracies in sensing information and pre-emption procedures based on data packet priorities to avoid interference.

## II. NR V2X Physical Layer Frame Structure

### A. Resource pool

In this chapter, the physical layer transmission structure of NR V2X is introduced. In NR V2X, resource pool (RP)s consisting of specific time and frequency resources can be (pre-)configured for SL transmissions [10]. A RP can be composed of contiguous sub-channels in the frequency domain and non-contiguous slots in

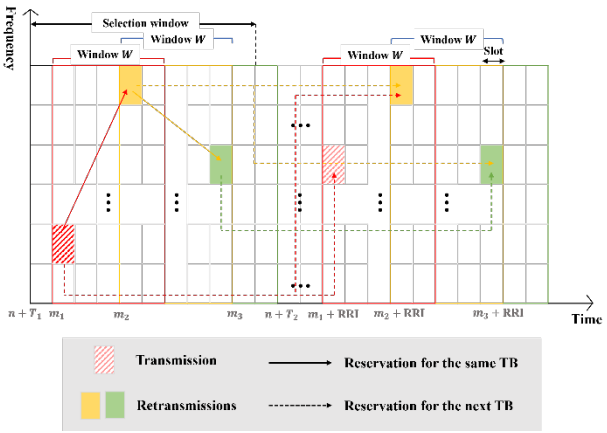



Fig. 2. Process for dynamic and semi-persistent PSSCH scheduling.

the time domain [10]. Here, sub-channels and slots refer to the minimum frequency and time resource units that can be used for SL transmissions. The RP can repeat every 10,240ms.

A single sub-channel can be comprised of multiple contiguous Physical Resource Blocks (PRB)s, and it can be configured with one value from {10, 12, 15, 20, 25, 50, 75, 100} PRBs [11]. Additionally, one PRB is defined as 12 consecutive subcarriers [12].

A single slot is composed of 12 or 14 Orthogonal Frequency Division Multiplexing (OFDM) symbols, depending on the length of the Cyclic Prefix (CP) [12]. Moreover, the starting position of SL symbols and the number of consecutive symbols transmitted within each slot are (pre-)configured within each RP. The number of consecutive symbols transmitted within a slot can vary from 7 to 14 symbols depending on the physical channel used for SL transmissions [10].

NR V2X supports various subcarrier spacing (SCS): {15kHz, 30kHz, 60kHz, 120kHz}. The length of slots in the time domain can be adjusted dynamically according to each SCS, such as {1ms, 0.5ms, 0.25ms, 0.125ms}[12].

### *B. Frame structure and scheduling for PSCCH/PSSCH*

In NR V2X communication, several channels are designated to facilitate SL transmissions. These channels encompass control channels, data channels, synchronization channels, and feedback channels [12]. This research is focusing on the Physical Sidelink Shared Channel (PSSCH) for sending data payloads and the Physical Sidelink Control Channel (PSCCH) for transmits control information. The PSCCH is used for the control data transmission to PSSCH scheduling and is responsible for transmitting the 1st-stage Sidelink Control Information (SCI) format utilized in NR V2X [12]. As shown in Fig. 1, PSCCH can be mapped to 2 or 3 OFDM symbols in the time domain and can be mapped to multiple PRBs within a single sub-channel in the frequency domain [12]. PSCCH is transmitted in the same slot associated with the corresponding PSSCH. In the frequency domain, it is mapped to the sub-channel with the lowest index among the scheduled sub-channels for PSSCH [13]. Therefore, the device can receive PSCCH by performing blind decoding for PSCCH candidate resources in all slots and sub-channels set as the RP. SCI format 1-A transmitted via PSCCH can include information such as priority, frequency RA information, time RA information, Resource Reservation Period (RRP), Demodulation Reference Signal (DMRS) pattern, 2nd-stage SCI format type, beta offset, DMRS port count, MCS (Modulation and Coding Scheme), additional MCS table indicators, Physical Sidelink Feedback Channel) PSFCH overhead indicators, and more [14].

PSSCH is mapped to one slot in the time domain and to multiple sub-channels in the frequency domain. In the time domain, the first symbol, corresponding to the start position of the SL symbols, is used for Adaptive Gain Control (AGC), the last symbol is used for the guard interval, and PSSCH can be mapped to the remaining symbols [10]. In the frequency domain, it can be mapped to the sub-channels of consecutive $L_{PSSCH}$ based on the frequency RA information indicated by SCI format 1-A, starting with the sub-channel where the PSCCH scheduling PSSCH is transmitted.

Fig. 2 illustrates the scheduling scheme for PSSCH in NR V2X. NR V2X supports both dynamic scheduling for non-periodic data traffic transmission and semi-persistent scheduling for periodic data traffic transmission [10].

In the case of dynamic scheduling, new resources are selected for a single transport block (TB), and scheduling information for PSSCH is dynamically conveyed to the device through the 1st-stage SCI format. For the same TB, a maximum of 2 blind re-transmissions is allowed, and for this purpose, up to 3 resources can be (pre-)allocated, including the initial transmission. Scheduling information, including the initial transmission and re-transmissions, is all delivered through the 1st-stage SCI format. In this case, the 1st-stage SCI is transmitted in the first slot of window $W$, which occupies 32 slots, and it conveys resource reservation information for re-transmissions within the range of $W$.

In the case of semi-persistent scheduling, multiple transmission resources can be (pre-)allocated according to a specific periodicity known as RRP (Resource Reservation Period) for the transmission of multiple different TBs. Possible RRP values for semi-persistent scheduling are {0, [1~99], 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000}ms and can be (pre-)configured in the RP. When selecting SL resources for periodic data traffic, the device can choose the RRP from the configured list that best suits the characteristics of the traffic to be transmitted. The selected RRP value is conveyed to the receiving device through the 1st-stage SCI format.

## III. NR V2X Resource Allocation Mode 2

In this chapter, the specific physical layer procedures for RA Mode 2 are introduced [10]. First, the high layer configuration parameters required for Mode 2 operation for RA are presented, and then the RA process is explained.

### *A. Parameter setting*

When operating in RA Mode 2, the high layer can request the UE to select resources for transmitting PSCCH and PSSCH. To facilitate the Mode 2 based RA procedure, various parameters, such as L1 priority ($p_{TX}$), Packet Delay Budget (PDB), the number of sub-channels used for PSCCH/PSSCH transmission, RRP ($RRP_{TX}$) for non-periodic scheduling, slot index set for re-evaluation, and slot index set for pre-emption, can be provided by the high layer.

When the RA procedure is triggered at slot $n$, the UE can define sensing and selection windows, as shown in Fig. 3. Within the sensing window, the UE can measure sensing values

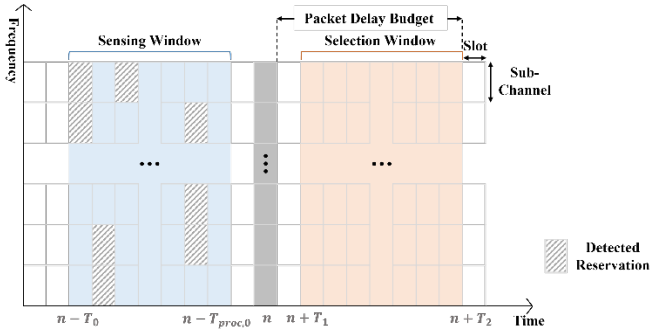



Fig. 3. Selection window and sensing window.

and based on these measurements, select transmission resources for PSCCH and PSSCH from within the selection window. The sensing window has a time range of $[n-T_0, n-T_{proc,0}]$. $T_0$ can be set to either 100ms or 1100ms, depending on the SCS. $T_{proc,0}$ represents the processing time required for sensing completion and can take values of {1, 2, 4} depending on the SCS.

The selection window spans $[n+T_1, n+T_2]$, where $T_1$ is the processing time required for the UE to select SL resources for identifying candidate resources before transmission. $T_1$ is determined within the range $0 \le T_1 \le T_{proc,1}$, and $T_{proc,0}$ can take values of $\{3, 5, 9, 17\}$. $T_2$ is determined based on the UE's implementation and falls within the range $T_{2min} \le T_2 \le PDB$. $PDB$ represents the maximum time limit for the transmission block to be transmitted, and $T_{2min}$ is composed of $\{1, 5, 10, 20\} \cdot 2^{\mu}(slots)$ms, depending on the SCS.

### *B. Resource sensing and selection procedure*

The UE first performs sensing for all resources within the sensing window range. To do this, the UE monitors all candidate resources within the sensing window where PSCCH can be transmitted. Through blind decoding, if PSCCH is detected in a specific PSCCH candidate resource, the UE can obtain allocation information in the time and frequency domains for PSSCH and reservation information for future transmissions from the corresponding SCI. The UE utilizes this information to check whether other UEs occupy transmission resources. If another UE is occupying the resources, the UE can measure the Reference Signal Received Power (RSRP) for the PSCCH's DMRS or the scheduled PSSCH DMRS. The decision of whether to use PSCCH's DMRS or PSSCH's DMRS for RSRP measurement can be preconfigured through high layer signaling. The measured RSRP values are then used in the subsequent resource selection process to determine whether resources should be excluded based on a comparison with RSRP threshold values.

The UE conducts sensing for all resources within the sensing window range. During this process, the UE decodes the 1st-stage SCI to distinguish between resources that are occupied or reserved by other UEs and candidate resources that can be used for future TBs. Decoding the SCI allows the measurement of the RSRP for the corresponding PSCCH or PSSCH. This information is later utilized in the resource exclusion and selection stages.

Based on the resource sensing results mentioned above, the UE can perform resource selection. First, the UE identifies candidates for all resources available for PSCCH and PSSCH transmissions within the selection window. Candidate resources are defined as a sequence of slots and continuous subchannels with $L_{PSSCH}$ options. $L_{PSSCH}$ falls within the range of $1 \le L_{PSSCH} \le L_{PSSCH,max}$, where $L_{PSSCH,max}$ represents the number of available subchannels for each slot and can be adjusted by a congestion control process. The UE excludes some resources from the identified candidate resource sets based on the following process.

First, exclude resources related to slots in which sensing could not be performed within the sensing window from the candidate resource set. The UE operates in a half-duplex mode, where it performs only one of the transmission or reception operations at a given time. In half-duplex operation, bidirectional communication is not possible. Therefore, for slots in which the UE was performing SL transmission within the sensing window, the UE cannot perform sensing. Assuming that the UE was transmitting in slot $s_i$ within the sensing window, the UE cannot detect the resources reserved by other UEs in slot $s_i$, and it excludes candidate resources corresponding to $s_i + q * RRP_k$ in the selection window. $RRP_k$ corresponds to all allowed RRP values based on the (pre-)configured RRP list for that transmission pool. $q$ is an integer in the range of $1 \le q \le Q$, where $Q$ represents the estimated number of periodic transmissions from the UE for each RRP. If the transmitting UE intends to perform semi-persistent scheduling, it additionally removes all candidate resources corresponding to slots $s_i + q * RRP_k$ in the selection window and slots $s_j$ corresponding to $s_j + j * RRP_{TX}$ in the selection window. $s_j$ represents the slot in which the UE intends to transmit within the selection window, and $1 \le j \le 10 * ReselectionCounter - 1$.

Second, the UE, during the sensing phase, compares the measured RSRP with (pre-)configured threshold values. If the measured RSRP is greater than the threshold value, the resources where interference signals occur are excluded from the candidate resources. RSRP threshold values are provided by the higher layer and can be set to different values depending on combinations of priority values. For example, if $p_i$ represents the priority for TB transmission from another UE received via the 1st-stage SCI, and $p_j$ is the priority of the transmitting UE selecting SL resources, different RSRP threshold values can be set for each $(p_i, p_j)$ combination. The transmitting UE, based on its own L1 priority value $p_{TX}$ for the traffic it intends to transmit and the priority value $p_i$ received from other UEs via the 1st-stage SCI during the sensing phase, can select an appropriate value from the RSRP threshold list to compare. The UE excludes candidate resources with RSRP values greater than the RSRP threshold $(p_i, p_j)$ in the selection window based on the 1st-stage SCI received from other UEs. If the UE receives another UE's 1st-stage SCI in slot $s_i$ of the sensing window, confirms the resource occupation period value as $RRP_{RX}$, and the measured RSRP value is greater than the threshold, the UE excludes all overlapping candidate resources at $s_i + q * RRP_{RX}$ positions in the selection window. If the transmitting UE intends to perform semi-persistent scheduling, it additionally excludes all candidate resources that overlap with the resources of other UEs and $s_j + j * RRP_{TX}$ slots mentioned earlier. After excluding candidate resources through the above process, the

UE checks whether the remaining candidate resources in the selection window are equal to or greater than $X\%$ of the total available candidate resources. If the proportion of remaining candidate resources in the selection window is lower than $X\%$, the RSRP threshold is increased by 3dB, and the RSRP comparison process can be performed again. In other words, if too many resources have been excluded, the threshold can be raised to reduce the number of excluded resources, thereby adjusting to make SL transmission more effective. $X$ can be set to one of the values {20, 35, 50} % in advance.

### C. Reselection procedure

When examining the resource sensing and selection process described above, it can be observed that there is a specific time gap between the moment of sensing and the moment of selecting transmission resources. If, after the transmitting UE has completed sensing, unexpected transmissions from other UEs occur, there is no guarantee that the resources chosen by the transmitting UE will remain valid. To address this issue, NR V2X has introduced a re-evaluation process that continuously performs sensing even after transmission resources have been selected through the above procedure to confirm the validity of the chosen resources.

When the UE initially selects transmission resources in slot m, it continuously performs resource sensing starting at least from $m - T_3$ slots earlier. $T_3$ is the maximum time required to perform the resource selection procedure. If the UE performs the resource sensing process again at slot $n'$, the UE can redefine a new selection window within the range $[n' + T_1, n' + T_2']$. Here, $T_2'$ can be defined as $T_{2min} \leq T_2' \leq PDB - (n' - n)$. The UE can reevaluate the availability of the resources selected at slot m, and if they are deemed unsuitable, it can choose transmission resources from the remaining candidate resources within the new selection window, excluding slot $m$. The re-evaluation process can be applied to both dynamic scheduling and semi-persistent scheduling cases.

### D. Pre-emption procedure

In RA Mode 2, a pre-emption feature has been introduced to differentiate the transmission order based on data traffic priorities. Thresholds for each priority are (pre-)configured in the RP. If a UE has occupied a specific resource but a UE with a higher priority wants to use the pre-empted resource, the pre-empting UE relinquishes the resource. This allows for more effective resource management when different priority traffic exists. Pre-emption is applicable to both dynamic scheduling and semi-persistent scheduling.

A UE can only become aware of another UE's usage of its reserved resources when it performs the resource sensing process again. Therefore, the resource sensing process for pre-emption occurs after the resource selection process. For example, if a transmitting UE completes the initial resource reservation in slot m, it can perform resource sensing for pre-emption after $m - T_3$ slots. When resource sensing for pre-emption is triggered at slot $n'$, the UE redefines a selection window within the range $[n' + T_1, n' + T_2]$. When a transmitting UE reserves resources, it checks pre-emption conditions to determine if higher priority UEs are using the resources. If another UE has a higher priority, the UE that reserved the resource relinquishes it and repeats the procedure to select SL resources within the newly defined selection window.

## IV. Conclusions

In this paper, we have introduced standard procedures for the physical layer of UEs in NR V2X for RA Mode 2. UEs operating in RA Mode 2 can monitor resources within the sensing window to determine the resource occupation status of other UEs. Based on sensing results, they can effectively select resources within the selection window for transmission. Furthermore, to address unexpected resource conflicts and differentiate transmission priorities based on data traffic priorities, re-evaluation and pre-emption procedures can be performed. NR V2X is expected to play a central role in realizing 5G-based autonomous driving and next-generation intelligent traffic systems by effectively supporting vehicle-to-vehicle communication in areas where base station services are unavailable.

## Acknowledgment

This research was supported by Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) (P0012724, The Competency Development Program for Industry Specialist).

## References

[1] T. Kim, Y. Kim, M. Jung and H. Son, “Intelligent Partial Sensing based Autonomous Resource Allocation for NR V2X,” *IEEE Internet of Things J.*, to appear.

[2] S. Chen, J. Hu, Y. Shi, L. Zhao, and W. Li, “A vision of C-V2X: Technologies, field testing, and challenges with Chinese development,” *IEEE Internet Things J.*, vol. 7, no. 5, pp. 3872–3881, May 2020.

[3] C. Shin, E. Farag, H. Ryu, M. Zhou and Y. Kim, "Vehicle-to-Everything (V2X) Evolution From 4G to 5G in 3GPP: Focusing on Resource Allocation Aspects," *IEEE Access*, vol. 11, pp. 18689-18703, 2023.

[4] S. A. Ashraf, R. Blasco, H. Do, G. Fodor, C. Zhang, and W. Sun, 1258 “Supporting vehicle-to-everything services by 5G new radio release-16 systems,” *IEEE Commun. Standards Mag*., vol. 4, no. 1, pp. 26–32, 1260 Mar. 2020.

[5] M. Harounabadi, D. M. Soleymani, S. Bhadauria, M. Leyh, and E. Roth-Mandutz, “V2X in 3GPP standardization: NR sidelink in release16 and beyond,” *IEEE Commun. Standards Mag*., vol. 5, no. 1, pp. 12–21, Mar. 2021.

[6] “NR; study on NR vehicle-to-everything (V2X), V16.0.0, release 16,” 3GPP, Sophia Antipolis, France, Rep. 3GPP TR 38.885, Mar. 2019.

[7] M. 1262 H. C. Garcia et al., “A tutorial on 5G NR V2X communications,” *IEEE Commun. Surveys Tuts*., vol. 23, no. 3, pp. 1972–2026, 3rd Quart., 2021.

[8] “NR sidelink evolution,” 3GPP, Sophia Antipolis, France, *3GPP document TSG RAN Meeting #94-e*, 3GPP RP-213678, Dec. 2021.

[9] V. Todisco, S. Bartoletti, C. Campolo, A. Molinaro, A. O. Berthet, and A. Bazzi, “Performance analysis of sidelink 5G-V2X mode 2 through an open-source simulator,” *IEEE Access*, vol. 9, pp. 145648–145661, 2021.

[10] *TSG RAN; NR; Physical Layer Procedure for Data, V17.2.0, Release 17*, 3GPP Standard TS 38.214, Jun. 2022.

[11] *TSG RAN; NR; Radio Resource Control (RRC) Protocol Specification, V17.1.0, Release 17*, 3GPP Standard TS 38.331, Jun. 2022.

[12] *TSG RAN; NR; Physical Channels and Modulation, V17.2.0, Release 17*, 3GPP Standard TS 38.211, Jun. 2022.

[13] *TSG RAN; NR; Physical Layer Procedure for Control, V17.2.0, Release 17*, 3GPP Standard TS 38.213, Jun. 2022.

[14] *TSG RAN; NR; Multiplexing and Channel Coding, V17.2.0, Release 17*, 3GPP Standard TS 38.212, Jun. 2022.

# Assessment of Dyspraxia in Adults

Dabin Choi[1] and Eun Young Kim[1,2,*]
[1]*Department of ICT Convergence, Soonchunhyang University, Asan-si 31538, Republic of Korea*
[2]*Department of Occupational Therapy, Soonchunhyang University, Asan-si 31538, Republic of Korea*
**Contact: eykim@sch.ac.kr, phone +82-41 530 4724*

*Abstract*—**Developmental coordination disorder (DCD) is characterized as movement problems that can manifest from childhood to adulthood. However, there is currently a lack of assessment instruments aimed at detecting DCD in the adult population of Korea. The objective of this research was to translate the Adult Developmental Coordination Disorder/Dyspraxia Checklist (ADC) with a focus on cultural adaptation. The translation procedure encompassed forward translation, reverse translation, 1st expert committee, 1st cognitive interview, 2nd expert committee, and 2nd cognitive interview The participants involved in the cognitive interviews were aged between 17.8 and 30.7 years, with a total of nine individuals. Items that elicited misunderstanding from three or more participants or suggestion of modifications during the cognitive interviews were deliberated upon and revised until a consensus was achieved to ensure the equivalence with the original text. In the case of altered items, further cognitive interviews were conducted involving three participants who faced difficulties in comprehension or exhibited misunderstandings among the selected participants. The expert committee evaluated the appropriateness of the terms. This study introduced Korea's inaugural screening tool for evaluating DCD in adults. This assessment is adaptable to a web-based context through the utilization of Information and Communication Technology applications.**

## I. INTRODUCTION

Developmental Coordination Disorder (DCD), based on the DSM-5 criteria, is a neurodevelopmental disorder characterized with motor skill deficits that significantly impact daily activities and cannot be attributed to intellectual or neurological conditions [1]. These motor deficits are observed as immature motor coordination, resulting in clumsy or inaccurate movements in motor skills. Also referred to as dyspraxia or clumsy individuals, DCD has a prevalence rate of 5-6% [1,2]

Around 50-70% of children diagnosed with DCD continue to experience clumsiness and motor difficulties into adolescence and adulthood [3]. Characteristics of adolescents and adults with DCD include challenges in tasks requiring complex motor skills, such as driving, writing quickly and neatly, team activities, and tasks involving both gross and fine motor skills. Furthermore, they may have difficulty in establishing relationships with peers due to a lack of social skills [3-5]. Adults with suspected DCD often exhibit a lower quality of life in social, psychological, and environmental domains [6].

The Adult Developmental Coordination Disorders/Dyspraxia Checklist (ADC), a self-report screening tool developed for adolescents and adults with Developmental Coordination Disorder (DCD), demonstrates high reliability and validity [7]. It is utilized to identify DCD in young adults [8-10].

The ADC was designed to contain a broad range of challenges related to DCD, including motor function, daily living skills, attention, time and spatial organization abilities, and social skills [7]. Recently, cultural adaptations of the ADC have been made in Germany and Italy, providing self-report screening tools for DCD [11-12]. However, as of now, there has been no self-report screening tool available in Korea for evaluating DCD in adolescents and adults.

The overall aim of this study was to translate and culturally validate the ADC for use in Korea.

## II. METHOS

### *A. Instruments: Adult Developmental Coordination Disorders/Dyspraxia Checklist*

The Adult Developmental Coordination Disorders/Dyspraxia Checklist (ADC) is a self-report questionnaire that can identify DCD in adolescents and adults aged 17 and above [7]. It consists of a total of 40 items, organized into three subscales: A (10 items), B (10 items), and C (20 items). Subscale A is related to difficulties in motor performance experienced during childhood. Subscale B pertains to current difficulties experienced by the individual in motor performance. Subscale C is associated with the motor performance symptoms observed by others. Each item is responded to using a 4-point Likert scale (Never=0, Sometimes=1, Often=2, Always=3). Scores of 56 or higher on the total scale indicate a risk of DCD, while scores of 65 or higher suggest a possibility of DCD. However, a minimum of 17 points must be scored in Subscale A for the interpretation of the total score. The internal reliability coefficient of the ADC is .95, with Subscale A at .91, Subscale B at .87, and Subscale C at .90. The ADC was translated into Korean after obtaining consent from the original author.

### *B. Procedures and Participants*

The translation of the ADC followed the procedure of Beaton [13] First, the forward translation team comprised two bilingual translators, whose native language was Korean. Independently, they translated from English to Korean. One of the bilingual translators did not have a clinical background related to occupational therapy. Second, the backward-translation was performed by two bilingual translators without a clinical background in occupational therapy, who independently translated from Korean to English. These translators were not provided with any prior information about the ADC and translated independently. Third, the initial expert committee consisted of one forward translator, two occupational therapists (with clinical experience of 26 and 28 years), and one back-translator. In the initial expert committee, the original text, forward translation, and back-translation were compared. They ensured semantic, idiomatic, experiential, and conceptual equivalence and made revisions to areas with inconsistencies, resulting in the completion of the pre-test version of the Korean ADC.

In order to assess comprehension and identify any cultural differences related to the content of the pre-test version of the Korean ADC, a first cognitive interview was conducted following the approach by Willis [14].

Nine adolescents and adults (5 males and 4 females) participated in the cognitive interview. The recruited participants' ages ranged from 17.8 to 30.7 years (mean age: 23.0 years, standard deviation: 3.3). Participants were recruited through university communities and social networking sites (SNS). One participant had dyslexia. The inclusion criteria for participants were: (1) being aged 16 or older and (2) being a native Korean speaker.

Participants completed of the pre-test version of the Korean ADC and were then asked three questions for each item: (1) How did you understand this item? (2) What behaviours (when you were a child/currently) led you to choose this option? (3) If you feel the item needs changes to make it easier to understand, how would you like to modify it? The cognitive interviews were recorded. During the cognitive interviews, items that were misunderstood by three or more participants or were suggested for modification were discussed, and revisions were made to enhance comprehension. The revised items were then subjected to a second cognitive interview with three participants selected based on their difficulty in understanding or misunderstanding the items.

To assess the understanding of the term "coordination," responses were collected from 11 adults in response to the question, "Please write down what activities come to mind when you think of 'activities that require coordination (movement control).'"

## III. RESULTS AND DISCUSSION

In the translation process, items 4, 8, and 12 were culturally adapted. In the original text, item 4 mentioned "volleyball," which was modified to "피구" (dodgeball), considering that volleyball is not as widely played in Korea as it is in English-speaking cultures. Additionally, for item 8, "violin" was changed to "피아노" (piano) because the piano is more popular than the violin in Korean culture, as determined by the expert committee.

For item 12 in the original text, eating utensils did not include "chopsticks." However, the expert committee considered "chopsticks" as a relevant tool for Korean culture, given the importance of chopsticks in dining practices in Korea. Therefore, "젓가락" (chopsticks) was added to the Korean translation for item 12.

A total of 11 items (2, 10, 11, 13, 21, 23, 27, 31, 24, 36, 40) had suggestions or were misunderstood by three or more participants. However, the expert committee chose not to modify five of these items (2, 13, 27, 36, 40) because providing specific examples or contexts could alter the sensitivity of the items. The modified items, except for one participant, were appropriately understood by the participants.

One participant interpreted the term "coordination" in the context of cooperation. To investigate whether others also found the term "coordination" difficult to understand or if they interpreted it appropriately, an additional survey was conducted. Respondents provided suitable activity examples like soccer and volleyball.

The expert committee determined that the term "coordination" was generally understood and did not require modification, thus establishing the Korean version of the ADC.

In accordance with previous studies [11], options were added to items 2, 25, and 32. This was done to consider individuals who may not have learned or attempted each activity. For item 2, the option "Indicate if you have not learned to ride a bicycle" was added. For item 25, the option "Indicate if you have not learned to drive" was included. For item 32, the option "Indicate if you do not drive" was appended. These options were added to ensure that individuals who have not learned or attempted these activities could provide accurate responses.

## IV. CONCLUSIONS

Based on the cultural adaptation, the ADC was translated into Korean. This involved identifying and modifying items that were misunderstood or suggested for changes through cognitive interviews, thus presenting the Korean version of ADC. In the future, the Korean version of ADC can be established online and provide the screening algorithm though Information and Communication Technology application.

## REFERENCES

[1] American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders (5th ed.). American Psychiatric Association.

[2] Magalhães, L. C., Missiuna, C., & Wong, S. (2006). Terminology used in research reports of developmental coordination disorder. Developmental Medicine and Child Neurology, 48(11), 937-941. https://doi.org/10.1017/S0012162206002040

[3] Kirby, A., Sugden, D., Beveridge, S., & Edwards, L. (2008). Developmental co‐ordination disorder (DCD) in adolescents and adults in further and higher education. Journal of Research in Special Educational Needs, 8(3), 120-131. https://doi.org/10.1111/j.1471-3802.2008.00111.x

[4] Cousins, M., & Smyth, M. M. (2003). Developmental coordination impairments in adulthood. Human Movement Science, 22(4-5), 433-459. https://doi.org/10.1016/j.humov.2003.09.003

[5] Kirby, A. (2011). The adult dcd/dyspraxia checklist (ADC) instructions for use (Kirby & Rosenblum, 2008) revision of scoring (2011). Retrieved from 〈 https://dyspraxiaaction.files.wordpress.com/2015/10/the-adult-dcd-guidelines-2011.pdf〉.

[6] Engel-Yeger, B. (2020). The role of poor motor coordination in predicting adults' health related quality of life. Research in Developmental Disabilities, 103, 103686. https://doi.org/10.1016/j.ridd.2020.103686

[7] Kirby, A., Edwards, L., Sugden, D., & Rosenblum, S. (2010). The development and standardization of the adult developmental co-ordination disorders/dyspraxia checklist (ADC). Research in Developmental Disabilities, 31(1), 131-139. https://doi.org/10.1016/j.ridd.2009.08.010

[8] Du, W., Wilmut, K., & Barnett, A. L. (2015). Level walking in adults with and without developmental coordination disorder: An analysis of movement variability. Human Movement Science, 43, 9-14. https://doi.org/10.1016/j.humov.2015.06.010

[9] Hyde, C., Fuelscher, I., Buckthought, K., Enticott, P. G., Gitay, M. A., & Williams, J. (2014). Motor imagery is less efficient in adults with probable developmental coordination disorder: Evidence from the hand rotation task. Research in Developmental Disabilities, 35(11), 3062-3070. https://doi.org/10.1016/j.ridd.2014.07.042

[10] Hyde, C., Fuelscher, I., Williams, J., Lum, J. A., He, J., Barhoun, P., & Enticott, P. G. (2018). Corticospinal excitability during motor imagery is reduced in young adults with developmental coordination disorder. Research in Developmental Disabilities, 72, 214-224. https://doi.org/10.1016/j.ridd.2017.11.009

[11] Meachon, E. J., Beitz, C., Zemp, M., Wilmut, K., & Alpers, G. W. (2022a). The adult developmental coordination disorders/dyspraxia checklist–German: Adapted factor structure for the differentiation of DCD and ADHD. Research in Developmental Disabilities, 126, 104254. https://doi.org/10.32872/cpe.4165

[12] Zappullo, I., Conson, M., Baiano, C., Cecere, R., Raimo, G., & Kirby, A. (2023). The Relationships between Self-Reported Motor Functioning and Autistic Traits: The Italian Version of the Adult Developmental Coordination Disorders/Dyspraxia Checklist (ADC). International Journal of Environmental Research and Public Health, 20(2), 1101. https://doi.org/10.3390/ijerph20021101

[13] Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. Spine, 25(24), 3186-3191. https://doi.org/10.1007/s10804-011-9121-3

[14] Willis, G. B. (2005). Cognitive interviewing: A tool for improving questionnaire design. Sage Publications./

# A Study on the Use of Tricycles for physical activity and safe mobility of the Elderly population

Wan Gyun Lee, Ji Yea Kim[1], Jae Rin Kim[2], Su Min Park[2], Min Seo Jung[2], Sang A Kim[2], Seong A Lee[2].

[1]*Department of ICT convergence, The Graduate School, Soonchunhyang University, Asan, South Korea*

[2]*Department of Occupational Therapy, College of Medical Science, Soonchunhyang University, Asan, South Korea*

*Contact: stuy8227@naver.com, phone +82, 010-6862-6216

*Abstract* — **This study evaluated the use of tricycles to promote physical activity and safe mobility for the elderly. A survey was conducted with 27 elderly individuals in a dementia-friendly village, exploring satisfaction and investigating the pros, cons, and improvements regarding tricycle usage. The satisfaction rate was high, with 74.1% responding that it contributed to their health and 63% stating that tricycles could be used safely. This study emphasizes the need for tailored tricycles considering safety and accessibility.**

## I. INTRODUCTION

The current increase in the elderly population in society implies a transition to an aging society, and this is having social and economic impacts.[1] The increase in the elderly population is primarily attributed to advances in medical technology and improvement in the quality of life, resulting in an extended average lifespan.[2] Consequently, there is a growing concern for the health and safety of the elderly. Maintaining the health of the elderly and promoting physical activity is considered a key factor in enhancing the quality of life. Physical activity in the elderly has a positive impact on their health, and a lack of physical activity can lead to the onset of chronic illnesses and functional decline.[3] Consistent exercise can contribute to improving the quality of life for the elderly through muscle strengthening, enhanced flexibility, and improved balance and coordination. Safe mobility within the community is considered a crucial element in enhancing the social participation and quality of life of the elderly.[4,5] Safe mobility can provide opportunities for the elderly to visit friends, family, hospitals, stores, and engage in social activities, as well as access necessary services. Elderly individuals utilize various means of transportation for mobility, including walking, cycling, public transportation, taxis, and personal vehicles, to facilitate their movement, The choice of transportation method varies depending on factors such as health status, economic conditions, transportation convenience, and geographic factors.[6]

In recent times, tricycles have gained popularity among the elderly population, especially for those with physical limitations. Tricycles provide opportunities for both exercise and mobility, ensuring safe movement for seniors. In this context, the present study aimed to enhance the physical activity and ensure safe mobility of the elderly by examining the usage of commonly available tricycles in the market. The study focused on identifying areas for improvement and enhancements related to the use of these tricycles.

## II. METHOD

The subjects of this study were 27 elderly individuals residing in the dementia-friendly village of OO-dong, Asan City. Participants were selected based on their voluntary agreement to participate in the study and their willingness to engage. The study investigated the usage and areas for improvement of tricycles through a quantitative study involving a satisfaction survey of tricycle riders and a qualitative study based on their perspectives and experiences riding tricycles.

## Ⅲ. Results

### 1. Satisfaction after Using Tricycles

Satisfaction after using tricycles was assessed in terms of stability, health improvement, psychological aspects, and physical activity. Under the stability category, 17 individuals (63%) responded 'yes' to the statement "Tricycles could be used safely without risks," indicating a high level of satisfaction regarding safety and stability. "In response to the statement 'I felt it contributed to improving my health,' 20 individuals (74.1%) agreed, and to the question 'Using tricycles contributes to safe mobility,' 19 individuals (82.5%) responded 'yes.' Regarding the psychological aspect, to the question 'I experienced enjoyment during the use of tricycles,' 20 individuals (74.1%) responded 'yes.' [Table 1.]

[TABLE I]
SATISFACTION AFTER USING TRICYCLES

| Item | Number of Respondents | Response Rate |
|---|---|---|
| Tricycles could be used safely without risks | 17 | 63 |
| Felt it contributed to health improvement | 20 | 74.1 |

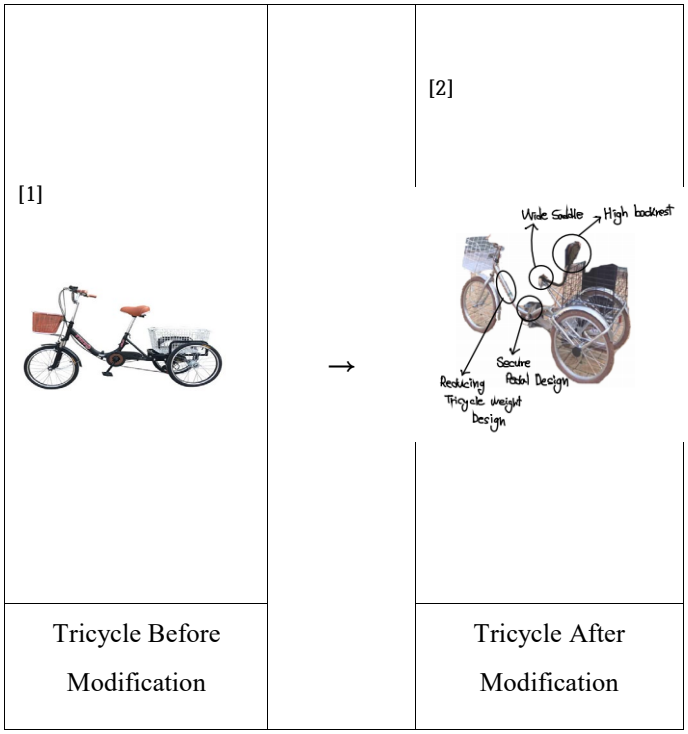
| | | |
|---|---|---|
| Tricycle usage contributes to safe mobility | 19 | 82.5 |
| Experienced enjoyment during the use of tricycles | 20 | 74.1 |

2. Qualitative Study on Tricycles

A qualitative study was conducted to investigate the perspectives and improvement suggestions regarding tricycle usage by the participants. The qualitative analysis revealed issues related to the weight of the tricycle, rotational problems due to the weight, stability of the foot on the pedal, and overall stability of the tricycle. Proposed solutions to address these issues were also identified. [Table 2.] Furthermore, in order to address these concerns and improvements, modifications and adjustments to the tricycle are planned as illustrated in [Figure 1.].

[Table 2.]
Opinions and Solutions after Using Tricycles

| Issues | Solutions |
|---|---|
| Weight and rotational issues due to tricycle weight | Reducing tricycle weight and efficient design |
| Stability of the foot on the pedal | Secure pedal design for foot stability |
| Overall stability of the tricycle | Wide saddle, backrest, and additional safety belt provided |




[Figure 1.]
Tricycle Before and After Modification.

The tricycle model is YHB-0020 by Dr. Bike company.[1] The tricycle after modification will have the following changes: Firstly, a digital dashboard will be attached to display kilometers, heart rate, and speed. Secondly, the overall tricycle structure will be altered to address issues related to weight. Thirdly, a wide saddle and high backrest will be provided to enhance stability.[2]

## III. Conclusions

This study aimed to assess the usability of tricycles designed for the elderly, considering economic feasibility and mass appeal. The research found positive responses in terms of satisfaction, stability, and both physical and psychological well-being when using tricycles among the elderly population. Tricycles were identified as a stable mode of transportation for the elderly. However, when examining tricycles available in the market, several issues and areas for improvement were identified, including problems related to weight, weight-induced bike rotation issues, pedal and footrest stability, and the stability of the tricycle saddle. These results indicate that tricycles offer a meaningful way for the elderly to engage in physical activity and ensure safe transportation.

Customized tricycle development tailored to the characteristics of the elderly population is necessary. Moreover, appropriate technology for enhancing physical activity and ensuring safe mobility for the elderly population is essential. However, it's important to note that there are limitations to this study, including the challenge of generalizing the results due to the specific elderly population studied in a particular region. Further research involving a more diverse elderly population is required to commercialize tricycles targeted at the elderly. Future research should focus on a more detailed analysis of the features of tricycles for the elderly, optimal bicycle structures, additional components, reliability assessments for application methods, and structural modifications.

## ACKNOWLEDGMENT

This research was supported by Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) (P0012724, The Competency Development Program for Industry Specialist).

## References

[1] United Nations, World Population Prospects 2019: Highlights, United Nations Department of Economic and Social Affairs, Population Division, 2019.

[2] J. W. Rowe & R. L. Kahn, "Successful aging," The gerontologist, vol. 37, no. 4, pp. 433-440, 1997.

[3] M. E. Nelson et al., "Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association," Circulation, vol. 116, no. 9, pp. 1094-1105, 2007.

[4] D. H. Paterson & D. E. Warburton, "Physical activity and functional limitations in older adults: a systematic review related to Canada's Physical Activity Guidelines," International Journal of Behavioral Nutrition and Physical Activity, vol. 7, no. 1, pp. 1-22, 2010.

[5] W. Li & T. H. Keegan, "Fast food and neighborhood stroke risk," The Annals of Family Medicine, vol. 8, no. 6, pp. 474-476, 2010.

[6] Q. L. Xue & L. P. Fried, "Chipping away at healthy aging: older adults' experience of urban change," The Gerontologist, vol. 38, no. 4, pp. 464-

473, 1998

# Facial Emotion Recognition based on Facial Points using Machine Learning Technique

*Vesal Khean[1,*], Ahsan Aziz[1], Chomyong Kim[2], Yunyoung Nam[3]*

*[1]Department of ICT Convergence, Soonchunhyang University, Asan, 31538, Republic of Korea*

*[2] ICT Convergence Research Centre, Soonchunhyang University, Asan, South Korea*

*[3] Department of Computer Science and Engineering, Soonchunhyang University, Asan, South Korea*

**Contact: kheanvisal1111@gmail.com*

*Abstract*— **Facial emotion recognition plays a pivotal role in human-computer interaction, affective computing, and various other domains. In this study, we propose an innovative approach to facial emotion recognition by leveraging facial key points extracted from images. We utilize machine learning methods, particularly the Support Vector Machine (SVM), to categorize facial emotion using these facial points. The dataset used for training and evaluation is derived from a CSV dataset collected through a virtual reality (VR) device. Our findings showcase the efficacy of the suggested approach, having attained a classification accuracy of 89% using the SVM model. This promising accuracy underscores the potential of using facial key points for emotion recognition, offering a non-invasive and efficient means of understanding human emotional states. This research contributes to the development of more natural and intuitive human-computer interfaces, as well as applications in fields such as healthcare, gaming, and user experience enhancement. Further exploration and refinement of this approach could lead to even more accurate and versatile facial emotion recognition.**

## I. INTRODUCTION

Facial expressions play a pivotal role in communication, encompassing both verbal and non-verbal forms of interaction. Non-verbal communication heavily relies on facial expressions, making Facial Expression Recognition (FER) a critical component of robotic vision. FER's primary goal is to accurately classify human facial expressions [1]. Identification of facial emotional expressions is crucial for everyday social functioning. Impairments in facial affect recognition have been found among patients with neurological and psychiatric disorders [2]. In recent years, FER has garnered substantial attention from researchers in the fields of computer vision and artificial intelligence, owing to its significant impact on both commercial and academic arenas. Its applications span a wide spectrum, including human-robot interaction, healthcare, augmented reality, road safety, virtual reality, deception detection, and surveillance.

Emotion detection techniques have evolved into three main categories: speech-based recognition, vision-based recognition, and hybrid approaches combining both visual and audio signals. Each approach has its unique advantages and limitations. However, hybrid techniques, leveraging multiple input types, tend to offer superior performance compared to single-input methods. Vision-based FER techniques utilize depth image sensors and RGB-based sensors [3]. Facial emotion recognition, a subfield of computer vision and affective computing, has gained significant attention in recent years due to its wide-ranging applications in human-computer interaction, healthcare, entertainment, and beyond. Understanding and accurately discerning human emotions from facial expressions are crucial for creating more intuitive and responsive computer systems. While this area of research has witnessed substantial progress, there remains room for improvement, particularly in achieving higher accuracy and robustness in emotion recognition.

In this proposed, we present a novel approach to facial emotion recognition based on the extraction and analysis of facial key points using machine learning techniques. Specifically, we employ Support Vector Machine (SVM), a powerful and well-established classification algorithm, to discern emotional states from facial features. The dataset used for training and evaluation in our research is sourced from a CSV dataset collected through a virtual reality (VR) device. This dataset captures a diverse range of facial expressions in a controlled yet dynamic environment, making it suitable for training and testing our proposed model.

The primary objective of this research is to investigate the potential of facial key points as informative cues for emotion recognition and to assess the effectiveness of SVM in this context. We aim to contribute to the development of more accurate and versatile systems for recognizing human emotions, which can have a profound impact on fields such as human-computer interaction, healthcare diagnostics, gaming, and user experience enhancement.

In this paper, we will first provide an overview of related work in facial emotion recognition and the role of facial key points in this domain. We will then describe our methodology, including data preprocessing, feature extraction, and the SVM-based classification approach. Subsequently, we will present and analyze the experimental results, highlighting the achieved accuracy and its implications. Finally, we will discuss the implications of our findings, potential applications, and avenues for future research in the realm of facial emotion recognition.

## II. RELATED WORK

FER models can be categorized into two main groups based on their approach to feature extraction: handcrafted and deep learning models. The FER task typically involves two key steps: feature extraction and classification. In the early stages, handcrafted models leveraged convolutional features such as curves, Local Binary Patterns (LBP), Haar wavelets, minutiae points, dynamic Bayesian networks (DBN), and edges. Subsequently, feedforward neural networks, Support Vector Machines (SVM), and extreme machine learning techniques were employed for classification [4].

In recent years, deep learning methods have significantly outperformed traditional handcrafted models in facial expression tasks. Carmen et al. [5] explored the impact of deep learning approaches on facial expression classification. In [6], a Deep Convolutional Neural Network (CNN) model with a SoftMax classifier was utilized for feature extraction and classification on the FER2013 dataset [7]. Orozco et al. [8] introduced transfer learning using popular architectures like AlexNet, VGG-19, and ResNet for FER tasks. Sun et al. [9] proposed a Deep CNN (DCNN) model, harnessing deep identity features for facial expression recognition. Barsoum et al. discussed FER using the VGG13 network on the FER+ dataset. For this task, they employed a weighted mixture DCNN with two channels. The first channel extracted facial features from grayscale images using a partially trained VGG16 model, while the second channel performed feature extraction using a shallow CNN on the CK+ and JAFFE datasets. Classification was carried out using the SoftMax function, combining the outputs from both channels through a weight fusion technique. Ruan et al. [10] introduced a feature decomposition and reconstruction method to effectively classify facial expression images.

Hua et al. [11] proposed three DCNN-based subnetworks trained independently on AffectNet and FER2013 datasets. These subnetworks were then ensembled using a weight fusion technique, achieving higher accuracy with larger datasets. Qing et al. [12] incorporated the few-shot learning concept into CNN for facial expression recognition in wild images. Wenmeng et al. [13] utilized a coattentive multi-task CNN for facial expression tasks, while Liu et al. [14] introduced a clip-aware expressive feature technique in CNN for video-based facial expression classification.

In [15], autoencoders were applied to merge geometric and regional LBP features, along with a Kohonen map, to perform the FER task. Identity-aware FER models have also been proposed over the years. Liu et al. [16] employed deep metric learning, optimizing a deep metric function and SoftMax classifier. Additionally, a deep comprehensive multi-patches aggregation network (DCMA-CNNs) was introduced [17], featuring two branches—one for holistic features and the other for local feature extraction from segmented image patches. These branches were later combined to classify facial expressions using a DCNN with ETI-pooling.

Finally, an ensemble of Multi-Level CNNs (MLCNNs) was proposed [18], which included a multi-level DCNN for feature extraction in facial expression images. Mohan et al. [19] introduced a gravitational force descriptor-based model for facial expression images, featuring two branches—one for geometric feature extraction and the other for holistic features extraction.

## III. Proposed Method

In this section, we delve into the intricacies of the proposed model employed for executing the Facial Expression Recognition (FER) task.

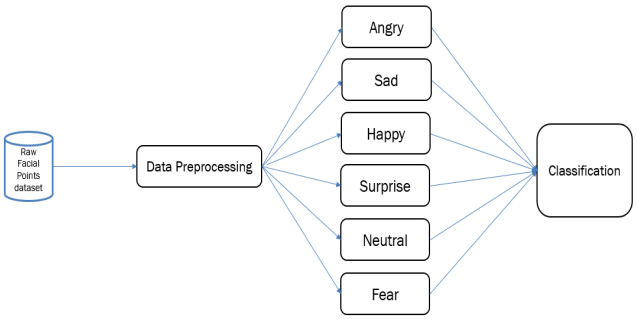



Fig. 1 The proposed whole flowchart

### A. Data Preprocessing

Labeling: Assign emotion labels to each data point in the dataset. These labels indicate the emotional expression associated with the corresponding set of facial key points. Accurate and consistent labeling is essential for supervised machine learning, as it provides the ground truth information that the model will learn to predict during training.

Data Scaling: This step is essential to ensure that all facial key points have a consistent scale, preventing certain key points from disproportionately influencing the machine learning model due to their numeric values. Scaling facilitates effective learning based on the relative positions of these key points, regardless of their initial numeric ranges, and is particularly important when using algorithms like Support Vector Machines (SVM).

### B. Machine Learning (ML)

In this step, machine learning, particularly the Support Vector Machine (SVM) model, is utilized for recognizing emotions from facial key points. The SVM model is trained to understand the connections between facial features and emotions, enabling it to make accurate emotion predictions based on new sets of facial key points. This approach enhances human-computer interaction and affective computing applications by enabling computers to understand and respond to human emotions effectively.

### C. Classification

Begin by preparing a balanced dataset with labelled facial key points. These points represent various expressions for each of the six emotions, happy, sad, angry, fear, surprise, and neutral. The key points are then preprocessed to ensure uniform scaling and prevent bias during model training. The SVM model is chosen for its effectiveness in multi-class classification tasks. It's trained on the preprocessed data to learn patterns in the facial key points associated with different emotions. Following training, the SVM model is evaluated using a separate test dataset to assess its classification accuracy.

## IV. Results and Discussion

### A. Dataset

The dataset collected from a Virtual Reality (VR) device is stored in the form of a CSV file, comprising 64 columns. Each row within this dataset encapsulates numeric values

representing facial key points, while the dataset encompasses six distinct emotion categories. This dataset provides a rich source of information for developing and training machine learning models to recognize and interpret facial expressions, offering valuable insights into human emotional states within immersive virtual environments. The column of the CSV file can be found in Figure 2.

| Facial Points in the Dataset | | | |
|---|---|---|---|
| BROW_LOWERER_L | BROW_LOWERER_R | INNER_BROW_RAISER_L | INNER_BROW_RAISER_R |
| CHEEK_PUFF_L | CHEEK_PUFF_R | JAW_DROP | JAW_SIDEWAYS_RIGHT |
| CHEEK_RAISER_L | CHEEK_RAISER_R | JAW_SIDEWAYS_LEFT | JAW_THRUST |
| CHEEK_SUCK_L | CHEEK_SUCK_R | LID_TIGHTENER_L | LID_TIGHTENER_R |
| CHIN_RAISER_B | CHIN_RAISER_T | LIP_CORNER_DEPRESSOR_L | LIP_CORNER_DEPRESSOR_R |
| DIMPLER_L | DIMPLER_R | LIP_CORNER_PULLER_L | LIP_CORNER_PULLER_R |
| EYES_CLOSED_L | EYES_CLOSED_R | LIP_FUNNELER_LB | LIP_FUNNELER_LT |
| EYES_LOOK_DOWN_L | EYES_LOOK_DOWN_R | LIP_FUNNELER_RB | LIP_FUNNELER_RT |
| EYES_LOOK_LEFT_L | EYES_LOOK_LEFT_R | LIP_PRESSOR_L | LIP_PRESSOR_R |
| EYES_LOOK_RIGHT_L | EYES_LOOK_RIGHT_R | LIP_PUCKER_L | LIP_PUCKER_R |
| EYES_LOOK_UP_L | EYES_LOOK_UP_R | LIP_STRETCHER_L | LIP_STRETCHER_R |
| LIP_SUCK_LB | LIP_SUCK_LT | MOUTH_LEFT | MOUTH_RIGHT |
| LIP_SUCK_RB | LIP_SUCK_RT | NOSE_WRINKLER_L | NOSE_WRINKLER_R |
| LIP_TIGHTENER_L | LIP_TIGHTENER_R | OUTER_BROW_RAISER_L | OUTER_BROW_RAISER_R |
| LIPS_TOWARD | LOWER_LIP_DEPRESSOR_L | LOWER_LIP_DEPRESSOR_R | UPPER_LID_RAISER_L |
| UPPER_LID_RAISER_R | UPPER_LIP_RAISER_L | UPPER_LIP_RAISER_R | |

Fig. 2 All the columns name of the dataset

The term "BROW_LOWERER_L" appears to refer to a specific facial muscle or muscle group that is involved in controlling the movement of the left eyebrow." CHEEK_RAISER_L" likely refers to the left side of the face that, when contracted, raise or lift the left cheek. Can be found in Figure 3.

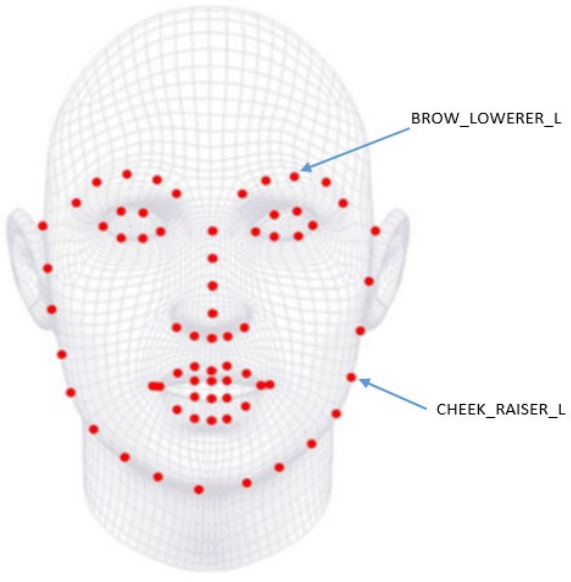



Fig. 3 Sample facial points represent to face

## B. *Experimentation settings*

The dataset used in this proposed was collected from a Virtual Reality (VR) device, capturing facial key points for analysis. A total of eight participants were involved in the data collection process, contributing to a diverse set of expressions. The dataset is structured into six distinct classes, each representing a unique emotional state. The experiments were conducted on a computing system equipped with 16GB of RAM and a 64-bit operating system, providing the necessary computational resources for data processing and machine learning tasks.

## C. *Experimental results*

In our experiments, we achieved an impressive accuracy rate of 89% when applying our machine learning model to the dataset collected from the VR device. This high level of accuracy demonstrates the model's effectiveness in recognizing and classifying emotions based on facial key points. Such results indicate promising prospects for employing this approach in various applications, including human-computer interaction. The result of confusion matrix can be found in Figure 4.

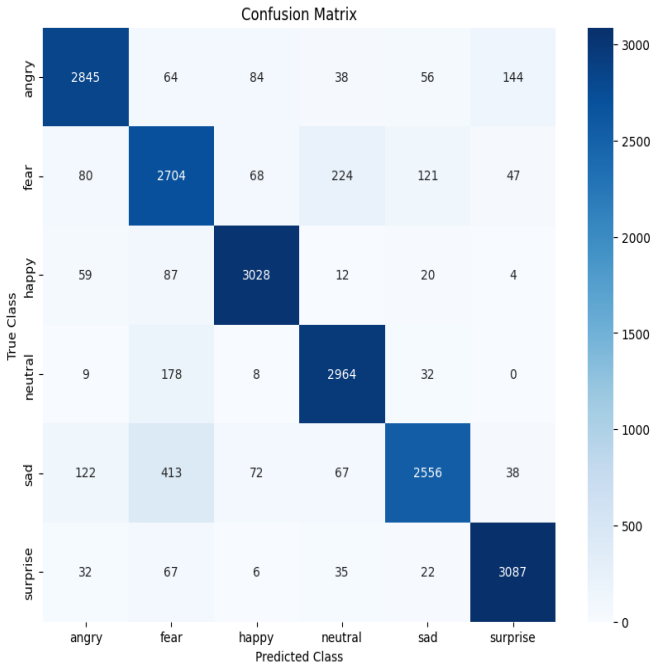



Fig. 4 Confusion Matrix of the proposed model on Test data

| Model | Accuracy % |
|---|---|
| Random Forest Classifier [17] | 70 |
| K-means [8] | 73 |
| KNN [18] | 70 |
| SVM [9] | 71 |
| Our proposed | 89 |

Fig. 5 Comparing the accuracy of our proposed model with state-of-the-art approaches

## V. CONCLUSION

In this paper, we applied the Machine Learning approach to the facial emotion from virtual reality dataset in CSV that token Support Vector Machine, demonstrating remarkably high recognition accuracy. Focused on facial emotion recognition using machine learning techniques, specifically utilizing Support Vector Machine (SVM) as the chosen model. The dataset employed in this research was sourced from virtual reality (VR) environments and was organized in a CSV format. One notable achievement of this proposed was the attainment of an impressive 89% accuracy rate in emotion recognition. This success demonstrates the efficacy of the SVM model in accurately classifying emotions based on facial key points extracted from the VR dataset.

Additionally, the high accuracy achieved through SVM showcases the model's robustness and its potential to enhance various applications, including human-computer interaction, affective computing, and virtual reality experiences. Moreover, the use of CSV-formatted data simplifies data management and integration into machine learning pipelines, making it accessible for researchers. Overall, this task serves as a valuable contribution to the field of emotion recognition, offering both methodological insights and practical applications in the realm of VR and human emotional understanding.

## ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ICAN(ICT Challenge and Advanced Network of HRD) program(IITP-2023-2020-0-01832) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation)

## REFERENCE

[1] Oloyede, Muhtahir O., and Gerhard P. Hancke. "Unimodal and multimodal biometric sensing systems: a review." IEEE access 4 (2016): 7532-7555.

[2] Geraets, C. N. W., S. Klein Tuente, B. P. Lestestuiver, M. Van Beilen, S. A. Nijman, J. B. C. Marsman, and W. Veling. "Virtual reality facial emotion recognition in social environments: An eye-tracking study." Internet interventions 25 (2021): 100432.

[3] Shan, Caifeng, Shaogang Gong, and Peter W. McOwan. "Facial expression recognition based on local binary patterns: A comprehensive study." Image and vision Computing 27, no. 6 (2009): 803-816.

[4] Liu, Ping, Shizhong Han, Zibo Meng, and Yan Tong. "Facial expression recognition via a boosted deep belief network." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1805-1812. 2014.

[5] Bisogni, Carmen, Aniello Castiglione, Sanoar Hossain, Fabio Narducci, and Saiyed Umer. "Impact of deep learning approaches on facial expression recognition in healthcare industries." IEEE Transactions on Industrial Informatics 18, no. 8 (2022): 5619-5627.

[6] Tang, Yichuan. "Deep learning using linear support vector machines." arXiv preprint arXiv:1306.0239 (2013).

[7] Goodfellow, Ian J., Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski et al. "Challenges in representation learning: A report on three machine

[8] Bekele, Esubalew, Zhi Zheng, Amy Swanson, Julie Crittendon, Zachary Warren, and Nilanjan Sarkar. "Understanding how adolescents with autism respond to facial expressions in virtual reality environments." IEEE transactions on visualization and computer graphics 19, no. 4 (2013): 711-720.

[9] Tabbaa, Luma, Ryan Searle, Saber Mirzaee Bafti, Md Moinul Hossain, Jittrapol Intarasisrisawat, Maxine Glancy, and Chee Siang Ang. "Vreed: Virtual reality emotion recognition dataset using eye tracking & physiological measures." Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies 5, no. 4 (2021): 1-20.

[10] Sun, Yi, Yuheng Chen, Xiaogang Wang, and Xiaoou Tang. "Deep learning face representation by joint identification-verification." Advances in neural information processing systems 27 (2014).

[11] Ruan, Delian, Yan Yan, Shenqi Lai, Zhenhua Chai, Chunhua

[12] Shen, and Hanzi Wang. "Feature decomposition and reconstruction learning for effective facial expression recognition." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 7660-7669. 2021.

[13] Hua, Wentao, Fei Dai, Liya Huang, Jian Xiong, and Guan Gui. "HERO: Human emotions recognition for realizing intelligent Internet of Things." IEEE Access 7 (2019): 24321-24332.

[14] Zhu, Qing, Qirong Mao, Hongjie Jia, Ocquaye Elias Nii Noi, and Juanjuan Tu. "Convolutional relation network for facial expression recognition in the wild with few-shot learning." Expert Systems with Applications 189 (2022): 116046.

[15] Yu, Wenmeng, and Hua Xu. "Co-attentive multi-task convolutional neural network for facial expression recognition." Pattern Recognition 123 (2022): 108401.

[16] Liu, Yuanyuan, Chuanxu Feng, Xiaohui Yuan, Lin Zhou, Wenbin Wang, Jie Qin, and Zhongwen Luo. "Clip-aware expressive feature learning for video-based facial expression recognition." Information Sciences 598 (2022): 182-195.

[17] Miranda, Catarina Runa, and Verónica Costa Orvalho. "Assessing Facial Expressions in Virtual Reality Environments." In VISIGRAPP (3: VISAPP), pp. 488-499. 2016.

[18] Zheng, Lim Jia, James Mountstephens, and Jason Teo. "Four-class emotion classification in virtual reality using pupillometry." Journal of Big Data 7 (2020): 1-9.

[19] Xie, Siyue, and Haifeng Hu. "Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks." IEEE Transactions on Multimedia 21, no. 1 (2018): 211-220.

[20] Nguyen, Hai-Duong, Sun-Hee Kim, Guee-Sang Lee, Hyung-Jeong Yang, In-Seop Na, and Soo-Hyung Kim. "Facial expression recognition using a temporal ensemble of multi-level convolutional neural networks." IEEE Transactions on Affective Computing 13, no. 1 (2019): 226-237.

[21] Mohan, Karnati, Ayan Seal, Ondrej Krejcar, and Anis Yazidi. "Facial expression recognition using local gravitational force descriptor-based deep convolution neural networks." IEEE Transactions on Instrumentation and Measurement 70 (2020): 1-12.

[22] Orozco, David, Christopher Lee, Yevgeny Arabadzhi, and Deval Gupta. "Transfer learning for facial expression recognition." Florida State Univ.: Tallahassee, FL, USA (2018).

# A study on predicting security vulnerabilities in graph embeddings

Myoungoh Choi[1*], Mincheol Shin[2], Hyunwook Yu[3] and Mucheol Kim[4]
[1]*Computer Science and Engineering, Chung-Ang University, Seoul, Korea*
[2]*Computer Science and Engineering, Chung-Ang University, Seoul, Korea*
[3]*Computer Science and Engineering, Chung-Ang University, Seoul, Korea*
[4]*Computer Science and Engineering, Chung-Ang University, Seoul, Korea*
*Contact: simwoo@cau.ac.kr, phone +82- 010 4438 6159

*Abstract*— **As the diversity of cyber security threats spreads, there is the disadvantage of taking a long time to prevent for security experts due to security vulnerabilities. Various studies are being conducted using deep learning models to identify attack techniques and predict security vulnerabilities.**

**Security vulnerability data exists in natural language form, and relationships between data may disappear during data processing, so a high level of expert knowledge is required.**

**This study presents a graph embedding-based security vulnerability prediction model using node2vec, which takes ideas from natural language processing. By presenting and considering various strategies utilizing CVE data, we suggest that the graph embedding method can be used to predict security vulnerabilities**.

## I. INTRODUCTION

A graph is a data structure that well represents various relationships in the real world and is made up of nodes and edges [1]. Through this, you can identify relationships between objects and extract unique characteristics from each graph.

Knowledge graphs, which structure and express knowledge in the real world, have the advantage of being easy to maintain consistency and accuracy of knowledge. It is used for information extraction and semantic analysis by connecting it with text data. Graph embedding has excellent advantages in terms of spatial efficiency and computational performance because it can be efficiently expressed as a low-dimensional vector while preserving the characteristics and relationships of the data [2]. These graph embedding models can be effectively used in various tasks such as clustering, classification, and link prediction [3].

Security vulnerabilities can expose users to many threats, including exposure of sensitive data and data modification in many systems, including operating systems, memory, and file system access. As cyber security threats have spread in recent years, there is a need for analysis of preventative measures against attacks that may occur in computer systems or software. Meanwhile, security vulnerability analysis to avoid cyber security threats has the disadvantage of taking a long time to prevent attacks because it is manually created by security experts. In a situation where security vulnerabilities are continuously increasing, it takes a lot of specialized knowledge and time to quickly respond to vulnerability attacks. In order to overcome these shortcomings, there are various studies to identify attack techniques and predict security vulnerabilities.

Security vulnerability data such as CVE contains various vulnerability information, but it is difficult to intuitively express the readability and relationships of the information. To compensate for these shortcomings, there have been recent studies predicting attack patterns, vulnerabilities, and software security vulnerabilities based on knowledge graphs [4, 5, 6]. These studies must be complex to effectively process and connect various data types such as CVE, CWE, and CAPEC through a knowledge graph, and a natural language processing process is required to understand and utilize text information [4]. Additionally, because data such as CAPEC and CVE data are not explicitly connected, direct connection is expensive and difficult [5].

In this study, data is expressed as a graph, the structural characteristics of the graph are identified through Node2vec, one of the graph embedding technologies, and security vulnerability characteristics and severity are predicted through natural language and connection patterns with nodes by adjusting parameters. Suggest whether it is possible. Graph embedding is effective in predicting similar relationships between data. By identifying similarities between security vulnerability data, vulnerabilities can be predicted more effectively. Existing studies focus on expressing knowledge in terms of relationships with entities, which can be understood from a structural perspective. Graph embedding explicitly represents connections between data, overcomes the difficulties of natural language processing, and improves the work efficiency of security vulnerability prediction. We aim to contribute to quick and effective security vulnerability prediction.

## II. REATED WORK

Due to the increase in security vulnerabilities, a variety of research is being conducted in the field of security vulnerabilities to predict vulnerabilities. Haddad et al. [7] used information from the National Vulnerability Database (NVD) to strengthen the analysis by mapping CVEs to hierarchical classes of the Common Weakness Enumeration (CWE). Han, Zhuobing, et al. [3] conduct research on security vulnerability prediction through knowledge graph representation learning and inference regarding software vulnerabilities and relationships. Das et al. [8] used CVE and CWE to identify vulnerabilities in computer systems and connect them to attack techniques, and Na et al. [9] classified CVE items into vulnerability types using a Naive Bayes classifier. There are ways to do it, etc.

Graph embedding techniques can be divided into Matrix Factorization Based Models, Random Walk Based Models, and Deep Neural Network Models [10]. Among them, representative algorithms based on Randomwalk include Deepwalk[11] and Node2vec[12].

Deepwalk walks between connected nodes in the network at the maximum length, constructs a walk with the discovered nodes, and determines similarity. When generating a walking sequence, a limited random walk is applied in the network to prevent the walking sequence from increasing infinitely. Afterwards, skip-gram, a word learning model, is applied to learn the node representation. For each walking sequence generated along Skip-gram, Deepwalk learns with the goal of maximizing the probability of a specific node's neighboring nodes.

Node2vec is an extension of DeepWalk and provides a more flexible random walk that can control the concept of node similarity. Node2vec is a method of learning and embedding relationships between neighboring nodes around a node, and can control the operation of a random walk through p and q. p is a parameter that controls the probability of returning from the current node to the previous node, and can be used to control whether the area surrounding the node is well searched. The larger the p value, the lower the probability of going to the visited node, and the smaller it is, the more it is generated only in nodes surrounding the starting node. q is a parameter that indicates whether a new node is well discovered from the starting node, and q can be used to control whether the random walk successfully discovers and explores new places. The random walk created by adjusting the parameters in this way consists of two stages: node sampling and skip-gram. To sample nodes, a node sequence is generated based on a random walk, and the skip-gram model is used to generate the node sequence considering that nodes in the same sequence have similar properties. and learn it.

## III. METHOD

In this section, we present a method for predicting security vulnerabilities through optimal graph size settings, hyperparameter settings, and neighbor node relationships using graph embedding technology.

Existing graph models have been mainly used for knowledge-based search and inference, but they need to be combined with other models for prediction. Node2Vec learns node embeddings by changing graph size, node relationships, hyperparameters, etc. according to the characteristics of graph data, and uses these embeddings as input to the prediction model..

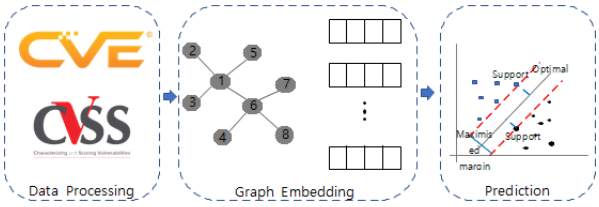



Fig. 1 Our Model

We balanced the model's efficiency, performance, and generalization ability by adjusting the size of nodes and edges. A large graph has more nodes and edges, which increases computation time and memory usage. We compare how the size of the nodes, edges, and classes of the graph affect the prediction of security vulnerabilities.

Node2Vec can understand and predict the similarity and relationship of security vulnerabilities by adjusting hyperparameters. If you set the p value to a large value, high weight is given to nodes similar to the current node, and similarity to surrounding nodes is emphasized. If you set the q value to a large value, relationships with various neighboring nodes are taken into consideration and the relationships of the entire graph are identified. By setting p and q, you can analyze and predict security vulnerabilities by identifying indirect relationships between other nodes or security vulnerabilities that are not directly connected to the security vulnerability.

p and q control the speed at which a walk searches for and leaves the neighbors of the starting node u. In particular, parameters allow interpolation between BFS and DFS, reflecting their affinity for different node equivalence concepts.

To predict security vulnerability classification, SVM [13] is used on the learned node embedding to perform node classification. SVM is one of the effective models for classification tasks. Since the boundary is defined considering the data distribution, nodes can be classified based on whether the graph embedding result matches the distribution of the data. It also works well with relatively small data. SVM is used to perform classification and prediction tasks in terms of graph size and relationship with nodes.

SVM uses a kernel trick to map samples into a high-dimensional space, which can be used to construct a hyperplane. This allows you to effectively handle binary and multi-class classification problems. We map the data to a high-dimensional space using the Gaussian kernel trick [14].

## IV. RESULT

Classification and evaluation were conducted with 46,139 data collected from 2016 to 2019 in CVE-Detail.

### A. Dataset

CVEID is the identification number of Common Vulnerabilities and Exposures (CVE) provided by NVD, CWEID is the identification number of software vulnerability information (Common WeaknessEnumeration, CWE) provided by MITER, and CVSS Scores (Common Vulnerability Scoring System) are vulnerability information. Severity score, Description is a detailed description of the vulnerability, Gained Access Level is how easily an attacker can access the system, Access is information about how an attacker can access the system through the vulnerability, and Complexity is information about how the vulnerability can be

accessed. The complexity of the attack to exploit, Authentication is the authentication requirement an attacker needs to perform the attack, Confidentiality is whether the vulnerability could allow an attacker to access confidential information, and Integrity is whether the vulnerability could allow an attacker to compromise data integrity. Availability is information about how an attacker could affect the availability of the system due to a vulnerability..

### B. Settings

To examine the performance of predicting security vulnerabilities for each node, a graph for each data was created. Each graph was created by dividing it into CVSS score, Description, CVSS Metrics (Gained Access Level, Access, Complexity, Authentication, Confidentiality, Integrity, Availability), and All.

CVSS is a graph representing the connections between nodes according to the severity of the vulnerability, Description is a graph representing the similarity between nodes based on the vulnerability description for the CVE-ID, CVSS Metrics is a graph with other columns as nodes, and All is a graph with all columns as nodes. It is a graph that goes through preprocessing and is expressed as a node.

We used Deepwalk and Node2vec to perform graph embedding. To find optimal performance in Deepwalk and Node2vec, the parameters were set as follows. The number of random walks and the length of the random walk were each set to 20, and the number of dimensions d was set to 128. Performance was optimized by adjusting $p \in \{0.5, 1, 2\}$ and $q \in \{0.5, 1, 2\}$. All experiments were repeated five times and the average values are reported in Table 1.

TABLE I
RESULTS OF 41,707 EXPERIMENTS FROM 2016 TO 2019 COLLECTED BY CVE-DETAIL

| | nodes | edges | classes | p | q | F1 |
|---|---|---|---|---|---|---|
| CVSS | 41,716 | 41,707 | 169 | 1 | 1 | 0.249 |
| | | | | 2 | 0.5 | **0.251** |
| | | | | 0.5 | 2 | 0.250 |
| Description | 93,852 | 618,508 | 169 | 1 | 1 | 0.697 |
| | | | | 2 | 0.5 | 0.696 |
| | | | | 0.5 | 2 | **0.699** |
| CVSS Metrics | 41,721 | 208,827 | 169 | 1 | 1 | **0.244** |
| | | | | 2 | 0.5 | 0.239 |
| | | | | 0.5 | 2 | 0.242 |
| All | 93,857 | 826,621 | 169 | 1 | 1 | 0.695 |
| | | | | 2 | 0.5 | 0.695 |
| | | | | 0.5 | 2 | **0.696** |

This study's performance evaluation uses accuracy and the F1 scale, which is the geometric mean of precision and recall. The formulas for accuracy and F1 scale are as equations (1) and (2), respectively.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (1)$$

$$F1\ Score = \frac{2}{\frac{1}{Precision}+\frac{1}{Recall}} = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (2)$$

### C. Results

The performance evaluation results for security vulnerability prediction are as follows.

When looking at the structural characteristics of the data, the graph was divided into data that explicitly listed vulnerabilities and data that was standardized by scoring.

In the case of structural characteristics, when comparing p and q values, there is no significant difference, ranging from 0.002 to 0.006. This is a result that well reflects structurally similar nodes and the connection pattern between nodes.

When comparing data that explicitly lists vulnerabilities and data that is standardized by scoring, F1 is about 0.69 for Description and ALL, which lists explicit data, and about 0.24 for CVSS Score and CVSS Metrics, so the difference in F1 between the two graphs is about 0.4. Be careful. Additionally, even when the number of classes was 169 and there were multiple classes, the graph with explicit data showed better performance overall.

## V. CONCLUSION

In this study, we proposed a security vulnerability classification prediction method combining Node2vec and SVM using software security vulnerability data, Common Vulnerabilities and Exposures (CVE). This method aims to construct a graph using each CVE data as a node, express each node as a low-dimensional vector using node embedding technology, and then classify and predict security vulnerabilities using SVM. Structural characteristics were compared in terms of connection patterns and parameter settings between nodes created according to the proposed node configuration method. Through this, we confirmed that our proposal can be used to automatically predict security vulnerability classification. The size and structure of the data, the connection pattern between nodes, and the composition of the graph affect the performance of the security vulnerability prediction model. We conducted experiments using various strategies to predict security vulnerabilities. Through the characteristics of the data, we were able to see that explicit data composed of natural language had excellent performance in multi-class classification and data based on standardized indicators. The significance of this study is to express explicit connections between data through graph embedding utilizing structural characteristics and to solve the difficulties of natural language processing. It was confirmed that security vulnerability data can be expressed in a graph form and that it is possible to classify and predict CWE (Common Weakness Enumeration), which is useful information for predicting vulnerabilities, using only the CVE ID and description. In this way, it is possible to overcome the existing limitations of requiring a lot of time and resources with professional knowledge, contributing to better understanding and response to quick and effective security vulnerability predictions, and can be used as an important tool in the field of cybersecurity.

## ACKNOWLEDGMENT

This research was supported by Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) (P0012724, The Competency Development Program for Industry Specialist).

## REFERENCES


[1] Chen, F., Wang, Y. C., Wang, B., and Kuo, C. C. J. (2020). Graph representation learning: a survey. APSIPA Transactions on Signal and Information Processing, 9, e15.

[2] Dourisboure, Yon, Filippo Geraci, and Marco Pellegrini. "Extraction and classification of dense communities in the web." Proceedings of the 16th international conference on World Wide Web. 2007.

[3] Xia, F., Sun, K., Yu, S., Aziz, A., Wan, L., Pan, S., and Liu, H. (2021). Graph learning: A survey. IEEE Transactions on Artificial Intelligence, 2(2), 109-127.

[4] Yuan, L., Bai, Y., Xing, Z., Chen, S., Li, X., & Deng, Z. (2021, July). Predicting entity relations across different security databases by using graph attention network. In *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)* (pp. 834-843). IEEE..

[5] Xiao, H., Xing, Z., Li, X., & Guo, H. (2019). Embedding and predicting software security entity relationships: A knowledge graph based approach. In Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part III 26 (pp. 50-63). Springer International Publishing.

[6] Han, Z., Li, X., Liu, H., Xing, Z., & Feng, Z. (2018, March). Deepweak: Reasoning common software weaknesses via knowledge graph embedding. In 2018 IEEE 25th International Conference on Software Analysis, Evolution and Reengineering (SANER) (pp. 456-466). IEEE.

[7] Haddad, A., Aaraj, N., Nakov, P., & Mare, S. F. (2023). Automated Mapping of CVE Vulnerability Records to MITRE CWE Weaknesses. arXiv preprint arXiv:2304.11130.

[8] Das, S. S., Serra, E., Halappanavar, M., Pothen, A., and Al-Shaer, E. (2021, October). V2w-bert: A framework for effective hierarchical multiclass classification of software vulnerabilities. In 2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA) (pp. 1-12). IEEE.

[9] Na, S., Kim, T., and Kim, H. (2017). A study on the classification of common vulnerabilities and exposures using naïve bayes. In Advances on Broad-Band Wireless Computing, Communication and Applications: Proceedings of the 11th International Conference On Broad-Band Wireless Computing, Communication and Applications (BWCCA–2016) November 5–7, 2016, Korea (pp. 657-662). Springer International Publishing.

[10] Hamilton, W. L., Ying, R., and Leskovec, J. (2017). Representation learning on graphs: Methods and applications. arXiv preprint arXiv:1709.05584.

[11] Perozzi, B., Al-Rfou, R., and Skiena, S. (2014, August). Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 701-710).

[12] Grover, A., and Leskovec, J. (2016, August). node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 855-864)

[13] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20, 273-297.

[14] Karimi, N., Kazem, S., Ahmadian, D., Adibi, H., & Ballestra, L. V. (2020). On a generalized Gaussian radial basis function: Analysis and applications. Engineering analysis with boundary elements, 112, 46-57.

# Under Blanket Sleep Posture Classification using RGB and Thermal Cameras based on Deep Learning Model

Awais Khan[1], Jung-Yeon Kim[2] and Yunyoung Nam[3*]

[1] *Department of ICT Convergence, Soonchunhyang University, Asan 31538, Republic of Korea*
[2] *ICT Convergence Research Center, Soonchunhyang University, Asan 31538, Korea*
[3] *Emotional and Intelligent Child Care System Convergence Research Center, Soonchunhyang University, Asan 31538, Republic of Korea*
[*] *Corresponding Author: Yunyoung Nam. Email: ynam@sch.ac.kr*

*Abstract*— **Sleep posture surveillance plays a crucial role in ensuring the well-being of bed-ridden patients and those vulnerable to falling out of bed. However, current sleep posture monitoring systems often struggle to account for the presence of blankets, limiting their practicality for comprehensive studies. To address this challenge, we propose a novel approach for sleep posture classification using RGB and thermal cameras, respectively. Our methodology begins by capturing a dataset of sleep postures using both the thermal and RGB cameras, encompassing videos of ten participants in six commonly observed sleep postures: supine, right log, left log, prone left, prone right and prone head. Initially, we normalized the dataset within video frames. Secondly, we fine-tune pre-trained VGG16 and ResNet50 models using transfer learning (TL). The features are then passed to the machine learning classifiers for the final classification. Notably, our dataset, obtained from Soonchunhyang University Asan, achieved an average accuracy of 96%.**

## I. INTRODUCTION

Optimal sleep is essential for maintaining both physical health and overall well-being [1]. Extensive research has established significant correlations between chronic conditions and sleep disorders like diabetes, obesity, and hypertension [2-4]. Poor sleep quality or disruption can exacerbate mental health concerns such as depression and anxiety [5]. Prevalence studies indicate that approximately 47 individuals per 1000 in the population are affected by sleep disorders, with some estimates suggesting an even higher prevalence [6]. In China, over 25% of adolescents are reported to experience sleep disturbances [7]. The most widespread sleep disorder is insomnia, which has been investigated through various methods including photoplethysmography, ballistocardiography, and polysomnography actigraphy [8].

Conventional approaches to measuring sleeping postures involve labour-intensive procedures such as manual video annotation or self-reported evaluations. However, these methods can be both time-consuming and prone to inaccuracies [9, 10] In more recent times, nonintrusive technologies have gained prominence in the realm of sleeping posture recognition. These technologies encompass a range of modalities including depth, infrared, visible light cameras [11, 12], wireless inertia measurement units, radar and radio sensors [13, 14], as well as pressure sensors [15, 16]. Despite these advancements, there remains a paucity of endeavors aimed at accurately recognizing and categorizing sleeping postures [17]. Addressing this gap, several machine-learning (ML) based strategies have developed, including approaches involving learning, k-nearest neighbors (KNNs), support vector machines (SVM), and convolutional neural networks (CNN). These techniques strive to enhance posture classification accuracy across both optical and pressure sensing modalities [18].

Polysomnography (PSG) stands as the definitive benchmark for sleep assessment, encompassing critical measures such as electroencephalography (EEG), electrooculography (EOG), electromyography (EMG), and supplementary data including breathing-related parameters and positional signals [17]. While PSG provides indispensable insights, its utilization entails notable drawbacks, including its high cost, logistical inconvenience, and time-intensive nature. This approach necessitates the attachment of multiple sensors to participants, a practice that could potentially influence the very sleep quality being studied. PSG data collection predominantly transpires within specialized sleep laboratories, environments characterized by high expenses and limited accessibility. Furthermore, the unfamiliar setting of a sleep laboratory might compromise sleep quality. Given these attributes, PSG encounters limitations in its applicability for prolonged monitoring, broad-scale implementation, or usage in populations averse to sensor attachment. The conventional polysomnography (PSG) procedure captures body position through a positioned sensor affixed to participants' abdomen using a belt. Nonetheless, the prevailing PSG scoring protocols, whether manual or automated, lack an exhaustive quantification of body position changes during sleep.

### *A. Major Challenges*

The key challenge to accurate noncontact monitoring, which has not yet been overcome, is tracking movements and postures through blankets, according to a state-of-the-art assessment of sleep surveillance technology. In order to overcome this problem, we set out to create a sleep position classifier utilizing deep learning models, an RGB and thermal camera, and with blankets.

### *B. Major Contributions*

The objective of this study is to address the limitations of current methods by introducing a novel deep learning approach framework for precise human sleep posture image classification. The proposed framework includes the following steps:

- The initial step involves converting the video dataset into individual frames.
- Modification of two pretrained deep learning models, VGG-16 and ResNet-50, by introducing an additional layer. This new layer establishes connections among the previous layers using fully connected (FC) layers.
- The features extracted from the modified models are employed for classification.

## II. METHODOLOGY

In this section, we propose a new deep learning method for human sleep poster image classification as illustrated in Figure 1. The methodology includes several distinct stages: initial pre-processing of the data, extraction of features utilizing pretrained models, and at last classification. This approach employs advanced techniques in deep transfer learning to enhance the performance of two existing pretrained models, namely VGG16 and ResNet-50. Subsequent to the extraction of features from these modified models, utilizing both RGB and thermal data, the feature vectors derived from the RGB and thermal data via both modified models are passed to the machine learning classifiers for the final classification. We proposed Cubic SVM, Linear SVM and Fine KNN for the classification of our proposed dataset.

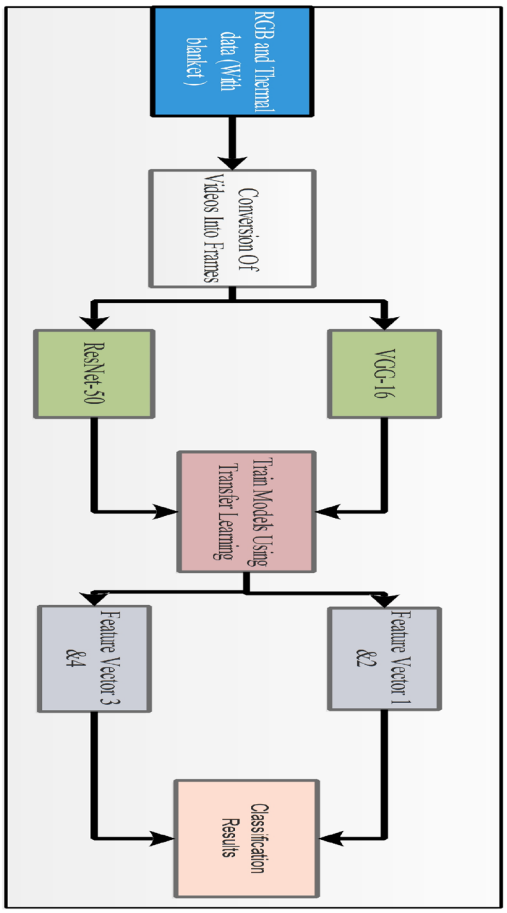



Figure 1. Proposed diagram of human sleep posture classification

### *A. Data Collection*

We engaged ten healthy adult males with an average age of 24 years. The participant had an average weight and height of 60.6 kg and 177 cm. All the participants were selected from Soonchunhyang University in Asan, South Korea. These people had no known histories of sleep disorders, musculoskeletal pain, deformities, or extreme sleep deprivation. Participants were positioned on a typical bed with dimensions of 196 cm in length, 90 cm in breadth, and 55 cm in height while data on sleeping positions were gathered. For the purpose of gathering data, we used both a thermal camera and an RGB camera. The RGB camera's resolution was 464 x 848 pixels, while the thermal camera's resolution was 624 x 832 pixels, and it functioned at a frame rate of six frames per second (fps). The data collection was performed by using the blanket in the following order: (1) supine, (2) prone left, (3) prone right, (4) prone head, (5) left log, and (6) right log. The sample of dataset is shown in Figure 2.

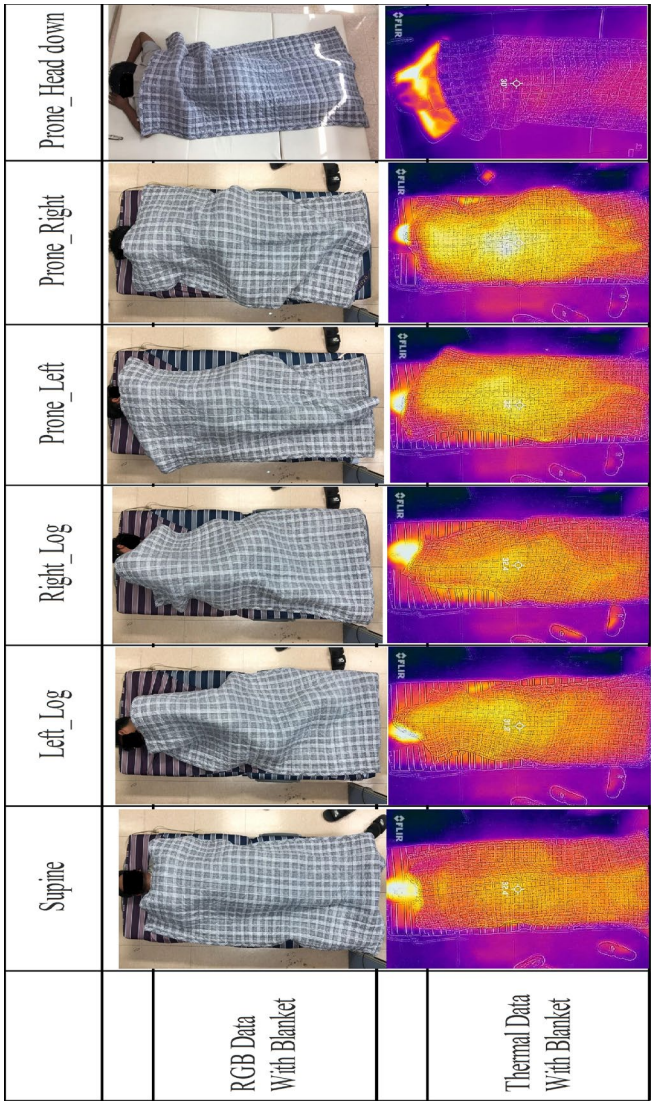



Figure 2. Sample of proposed dataset.

### B. Modified VGG16

The notable feature of the VGG16 model is its focus on simplicity [55], characterized by the use of uniform $2 \times 2$ max-pooling layers (PL) and $2 \times 2$ max-pooling layers (MPL) with a stride of $2$, along with $3 \times 3$ filter convolutional layers at stride $1$. The model employs a sequence of convolutional and pooling layers, followed by fully connected layers. With a total of $16$ layers, comprising $13$ convolutional and three fully connected layers, the architecture is illustrated in Figure 3. Originally trained on the ImageNet dataset with $224 \times 224 \times 3$ input dimensions, we adapt the model by omitting the last fully connected layer and adding a new one with six number of classes: supine, left log, right log, prone head, prone left, and prone right. This modified model is fine-tuned using transfer learning on the specified human sleep poster image data. Feature extraction is performed from the last fully connected layer, yielding a vector of dimension $N \times 4096$, where the last layer's output is $N \times 6$.

### C. Modified Resnet50

The ResNet architecture has gathered appreciation for its exceptional performance in facilitating a more direct information flow within the network, effectively addressing the issue of disappearing gradients during backpropagation. Thus, the ResNet is well-regarded for its efficient handling of deep networks. It addresses the vanishing gradient problem through shortcut connections. ResNet-50, with its 23 million parameters, was adapted for our sleep posture study. We replaced the final classification layer and used advanced transfer learning techniques to fine-tune the model. This customized model excels at feature extraction, yielding N × 2048-dimensional feature vectors.

## III. Results And Discussion

In this section, we have conducted the detailed experimental process of the proposed framework. The results are presented using visual graphs and well-defined performance measures to provide a comprehensive and clear valuation of our methodology's. In this study, the HSP (Human Sleep Posture) image dataset was divided into training and testing sets with 80:20 ratio. The training process was configured with specific parameters, including 100 iterations, 100 epochs, a minibatch size of 34, and a learning rate set at 0.0001. Stochastic Gradient Descent (SGD) served as the optimization algorithm. A 5-fold cross-validation was executed, assessing multiple classifiers across a range of performance metrics, including precision, rate, recall, and accuracy. All simulations were conducted using MATLAB 2022a. The study was carried out on a Core i7 processor and 8 GB of RAM.

### C. Experimental results with blanket using RGB image dataset

The classification results of the VGG16 and ResNet50 model for the HSP dataset using thermal images and with blankets, are shown in Table I. The extracted test features from VGG16 and ResNet50 were passed to three ML classifiers. The highest accuracy of 95.8% was achieved by the LSVM classifier when using the VGG16 model. This model shows performance metrics, including computational time, recall rate and precision rate, values of 76 (Sec), 95.6 and 95.5, respectively. The second highest accuracy of 95.4% was obtained by CSVM classifier, resulting in computational time, recall rate, precision rate, and AUC values of 76.2 seconds, 96 and 95.8. The highest accuracy of 95.1% was achieved by the Linear SVM classifier when the ResNet50 model was employed. This model shows performance metrics, including computational time, recall rate, precision rate, and AUC values of 25.7 (Sec), 95.3 and 95. Cubic SVM classifier got the second highest accuracy 94.9%, with corresponding performance values of 27 seconds, 94.9 and 94.9 respectively. Remarkably, the gap in accuracy between these two top-performing models was only 0.2%. The confusion matrix for the VGG16 and LSVM model is shown in Figure 3.

**Table I.** Proposed results VGG16 and ResNet50 model using the RGB image dataset with blanket.

| DL Models | ML Classifiers | Accuracy | Time (sec) | Precision Rate | Recall Rate |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| | 1- CSVM | 95.4% | 76.2 | 95.8 | 96 |
| **VGG16** | **2- LSVM** | **95.8%** | **76** | **95.5** | **95.6** |
| | 3- FKNN | 95.2% | 78.2 | 95.7 | 95.9 |
| | 1- CSVM | 94.9% | 27 | 94.9 | 94.9 |
| **ResNet-50** | **2- LSVM** | **95.1%** | **25.7** | **95** | **95.3** |
| | 3- FKNN | 94.8% | 39.1 | 94.7 | 94.7 |


**Figure 3.** Confusion matrix for RGB image classification with blanket using VGG16 model with 5-fold cross validation.

### *D. Experimental results with blanket using thermal image dataset*

The classification results of the VGG16 and ResNet50 model for the HSP dataset using thermal images and with blankets, are shown in Table II. The extracted test features from VGG16 and ResNet50 were passed to three ML classifiers. The highest accuracy of 96% was achieved by the LSVM classifier when using the VGG16 model. This model shows performance metrics, including computational time, recall rate and precision rate, values of 91.3 (Sec), 95.01, 95.2, and 0.98, respectively. The second highest accuracy of 94.9% was obtained by CSVM classifier, resulting in computational time, recall rate, precision rate, and AUC values of 85.3 seconds, 95 and 95.4. The highest accuracy of 94.5% was achieved by the Linear SVM classifier when the ResNet50 model was employed. This model shows performance metrics, including computational time, recall rate, precision rate, and AUC values of 28 (Sec), 94.7 and 94.5. Cubic SVM classifier got the second highest accuracy 94.4%, with corresponding performance values of 29.4 seconds, 94.3 and 94.3 respectively. Remarkably, the gap in accuracy between these two top-performing models was only 0.1%. The confusion matrix for the VGG16 and LSVM model is shown in Figure 4.

**Table II.** Proposed results VGG16 and ResNet50 model using the Thermal image dataset with blanket.

| DL Models | ML Classifiers | Accuracy | Time (sec) | Precision Rate | Recall Rate |
|---|---|---|---|---|---|
| | 1- CSVM | 94.9% | 85.3 | 95 | 95.4 |
| **VGG16** | **2- LSVM** | **96%** | **91.3** | **95.01** | **95.2** |
| | 3- FKNN | 94.2% | 138.1 | 94.2 | 94.6 |
| | 1- CSVM | 94.4% | 29.4 | 94.3 | 94.3 |
| **ResNet-50** | **2- LSVM** | **94.5%** | **28** | **94.5** | **94.7** |
| | 3- FKNN | 94.3% | 61.3 | 94.2 | 94.2 |


**Figure 4.** Confusion matrix for Thermal image classification with blanket using VGG16 model with 5-fold cross validation.

## IV. CONCLUSIONS

The goal of this research is to create a sleep surveillance system capable of continuously monitoring participants' sleeping postures and behaviours. This study marks a significant milestone in establishing foundational parameters for posture classification and addressing the practical challenges associated with blankets in a controlled environment. Moving forward, the next critical phase involves transitioning to real-world sleeping conditions and assessing the distribution of each posture over time. This progression will enable a more comprehensive understanding of sleep patterns in natural settings.

This study underscores the effectiveness of our fine-grained six-posture system, even when considering the potential impact of blankets, achieving a satisfactory overall

classification accuracy of 96%. These findings lay a solid foundation for the practical implementation of sleep surveillance systems in care homes and hospitals. Moving forward, we plan to conduct field tests to validate the system's performance in real-world settings, and we aim to diversify our dataset by including participants with various health and as well as different blanket conditions. Additionally, we will enhance the system's capabilities to not only classify postures but also identify body morphotypes, body segment positions, and joint angles, further advancing its utility in sleep monitoring and healthcare applications.

## ACKNOWLEDGMENT

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ICAN (ICT Challenge and Advanced Network of HRD) program (IITP-2023-2020-0-01832) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation).

## REFERENCES

[1] F. Lin *et al.*, "SleepSense: A noncontact and cost-effective sleep monitoring system," vol. 11, no. 1, pp. 189-202, 2016.

[2] M. Khalil, N. Power, E. Graham, S. S. Deschênes, N. J. D. r. Schmitz, and c. practice, "The association between sleep and diabetes outcomes–A systematic review," vol. 161, p. 108035, 2020.

[3] R. D. Vorona, M. P. Winn, T. W. Babineau, B. P. Eng, H. R. Feldman, and J. C. J. A. o. i. m. Ware, "Overweight and obese patients in a primary care population report less sleep than patients with a normal body mass index," vol. 165, no. 1, pp. 25-30, 2005.

[4] K. Spiegel, K. Knutson, R. Leproult, E. Tasali, and E. J. J. o. a. p. Van Cauter, "Sleep loss: a novel risk factor for insulin resistance and Type 2 diabetes," 2005.

[5] M. A. Short, S. A. Booth, O. Omar, L. Ostlundh, and T. J. S. m. r. Arora, "The relationship between sleep duration and mood in adolescents: A systematic review and meta-analysis," vol. 52, p. 101311, 2020.

[6] A. Hombali *et al.*, "Prevalence and correlates of sleep disorder symptoms in psychiatric disorders," vol. 279, pp. 116-122, 2019.

[7] M. Liang, L. Guo, J. Huo, and G. J. P. O. Zhou, "Prevalence of sleep disturbances in Chinese adolescents: A systematic review and meta-analysis," vol. 16, no. 3, p. e0247333, 2021.

[8] A. Y.-C. Tam, B. P.-H. So, T. T.-C. Chan, A. K.-Y. Cheung, D. W.-C. Wong, and J. C.-W. J. S. Cheung, "A blanket accommodative sleep posture classification system using an infrared depth camera: A deep learning approach with synthetic augmentation of blanket conditions," vol. 21, no. 16, p. 5553, 2021.

[9] D. Cary, K. Briffa, and L. J. B. o. McKenna, "Identifying relationships between sleep posture and non-specific spinal symptoms in adults: A scoping review," vol. 9, no. 6, p. e027633, 2019.

[10] T. Kubota *et al.*, "Characteristic features of the nocturnal sleeping posture of healthy men," vol. 1, no. 2, pp. 183-185, 2003.

[11] M. Yu, A. Rhuma, S. M. Naqvi, L. Wang, and J. J. I. t. o. i. t. i. b. Chambers, "A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment," vol. 16, no. 6, pp. 1274-1286, 2012.

[12] M. Masek, C. P. Lam, C. Tranthim-Fryer, B. Jansen, and K. J. S. Baptist, "Sleep monitor: A tool for monitoring and categorical scoring of lying position using 3D camera data," vol. 7, pp. 341-346, 2018.

[13] J. Liu, X. Chen, S. Chen, X. Liu, Y. Wang, and L. Chen, "TagSheet: Sleeping posture recognition with an unobtrusive passive tag matrix," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, 2019, pp. 874-882: IEEE.

[14] F. Zhang *et al.*, "SMARS: Sleep monitoring via ambient radio signals," vol. 20, no. 1, pp. 217-231, 2019.

[15] J. J. Liu *et al.*, "A dense pressure sensitive bedsheet design for unobtrusive sleep posture monitoring," in *2013 IEEE international conference on pervasive computing and communications (PerCom)*, 2013, pp. 207-215: IEEE.

[16] E. J. Pino, A. D. De la Paz, P. Aqueveque, J. A. Chávez, and A. A. Morán, "Contact pressure monitoring device for sleep studies," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 4160-4163: IEEE.

[17] K. Tang, A. Kumar, M. Nadeem, and I. J. I. Maaz, "CNN-based smart sleep posture recognition system," vol. 2, no. 1, pp. 119-139, 2021.

[18] G. Matar, J.-M. Lina, G. J. I. j. o. b. Kaddoum, and h. informatics, "Artificial neural network for in-bed posture classification using bed-sheet pressure sensors," vol. 24, no. 1, pp. 101-110, 2019.

# Immersive Emotion Recognition in Virtual Reality: A Machine Learning Approach with Facial Points

*Ahsan Aziz[1], Chomyong Kim[2], Yunyoung Nam[3]*

*[1]Department of ICT Convergence, Soonchunhyang University, Asan, 31538, Republic of Korea*

*[2] ICT Convergence Research Centre, Soonchunhyang University, Asan, South Korea*

*[3] Department of Computer Science and Engineering, Soonchunhyang University, Asan, South Korea*

**Contact: ynam@sch.ac.kr*

*Abstract*— **Facial emotion recognition plays an essential role in human-computer interaction, affective computing, and various other domains. In this study, we propose a state-of-the-art approach to facial emotion recognition we present a novel approach to facial emotion recognition the initial phase of our research involved the acquisition of emotional data utilizing a Virtual Reality (VR) device. We enlisted eight participants to contribute to this data collection process, tasking them with displaying predetermined emotions while exposed to a set of sample images. These sample images encompassed six distinct emotional classes, including happiness, sadness, anger, neutrality, surprise, and fear.**

**The VR device was instrumental in capturing and storing the data in CSV files, resulting in datasets exceeding 2000 rows and comprising 64 columns. Each column was meticulously categorized to correspond to specific facial movement and expression metrics. Subsequently, we diligently undertook data tagging and pre-processing, ensuring uniform dimensions across all files, a critical step in facilitating consistent and reliable analyses. our machine learning model, particularly the cubic Support Vector Machine (SVM), demonstrated exceptional performance. We achieved a remarkable accuracy rate of 99.1% when applying this model to the VR dataset for emotion recognition based on facial key points. This extraordinary level of accuracy underscores the robustness and efficacy of our approach. Such outstanding results hold great promise for the application of our methodology across a broad spectrum of fields, with particular relevance to the advancement of human-computer interaction.**

## I. INTRODUCTION

In an increasingly digitalized world, the fusion of technology and human emotion has emerged as a promising frontier, offering potential applications across diverse domains, including healthcare, education, entertainment, and human-computer interaction. Among the avenues of exploration, facial emotion recognition stands out as a potent tool for comprehending and enriching human experiences. The ability to decipher and respond to human emotions holds immense significance not only for creating more empathetic and responsive technology but also for its profound implications in fields such as psychology, marketing, and artificial intelligence.

Conventional methods of facial emotion recognition primarily rely on the analysis of images and videos, extracting insights from two-dimensional representations of facial expressions. Nevertheless, these approaches possess inherent limitations in capturing the intricacies and subtleties of human emotions, which often manifest through the three-dimensional dynamics of facial features. This is where Virtual Reality (VR) comes into play—a technology that has revolutionized the manner in which we engage with digital content and, more notably, offers a distinctive avenue for the observation and analysis of emotions in an immersive and genuine context. In [1] it investigates the utility of virtual reality (VR) in the assessment and training of emotion recognition within social contexts, utilizing realistic and dynamic stimuli and involves a comparative analysis of three emotion recognition tasks: VR, video, and photo tasks. This paper embarks on a journey into the captivating realm of facial emotion recognition, employing a Virtual Reality device equipped with facial point tracking technology in tandem with the capabilities of machine learning. By transcending the constraints of flat imagery and videos, this innovative approach unlocks fresh dimensions in the interpretation, understanding, and responsive handling of human emotions. This introductory section sets the stage for the exploratory voyage that follows, laying out the foundational concepts and driving forces underpinning this research.

In [2] the author acknowledges the distinctive nature of data collection within virtual reality (VR) environments, particularly through head-mounted displays, when contrasted with traditional classroom or online learning settings. This underscores the necessity for developing a recognition approach tailored specifically to VR contexts and they have a novel method for recognizing learning concentration within VR environments. Through the immersion of users into computer-generated environments, VR transcends the limitations of traditional screens, bestowing a sense of presence and embodiment that previously resided in the realm of science fiction. With the advent of increasingly sophisticated VR devices, boasting capabilities such as positional tracking, haptic feedback mechanisms, and precise motion capture, the potential for applications has expanded exponentially. In [3] the paper describes the creation of an inventive VR-based system designed for presenting facial emotional expressions. This system not only enables the monitoring of emotional expressions but also tracks eye gaze and physiological signals connected to the identification of emotions,

introducing more efficient therapeutic methodologies. One particularly promising avenue within the VR spectrum lies in its capacity to mirror and manipulate human emotions. By furnishing a platform capable of replicating real-world scenarios and interpersonal interactions, VR provides an ideal environment for eliciting and scrutinizing emotions under controlled and customizable conditions. This paper navigates through the utilization of VR as a conduit for capturing the intricate choreography of facial expressions in three dimensions, thereby amplifying our ability to identify and appreciate emotions in a more comprehensive manner.

Facial emotion recognition is an intricate undertaking, heavily reliant on the analysis of facial expressions composed of numerous facial landmarks, including the positions of the eyes, eyebrows, nose, and mouth. These landmarks play a pivotal role in conveying emotions and subtle cues that often elude detection without a high degree of precision.

In recent years, advancements in the domains of computer vision and machine learning have staged a revolution in the field of facial emotion recognition. Machine learning algorithms, particularly the formidable deep learning models, have exhibited remarkable prowess in autonomously extracting features from facial data, thus enabling the accurate recognition of emotions. This paper delves into the methodologies and techniques employed to harness the power of machine learning in deciphering the wealth of data gleaned from VR devices and facial point tracking.

In this proposed, we present a novel approach to facial emotion recognition the initial phase of our research involved the acquisition of emotional data utilizing a Virtual Reality (VR) device. We enlisted eight participants to contribute to this data collection process, tasking them with displaying predetermined emotions while exposed to a set of sample images. These sample images encompassed six distinct emotional classes, including happiness, sadness, anger, neutrality, surprise, and fear.

The VR device was instrumental in capturing and storing the data in CSV files, resulting in datasets exceeding 2000 rows and comprising 64 columns. Each column was meticulously labeled to correspond to specific facial movement and expression metrics. Subsequently, we diligently undertook data labeling and preprocessing, ensuring uniform dimensions across all files, a critical step in facilitating consistent and reliable analyses. Our research then proceeded to employ diverse machine learning algorithms, with a primary focus on Support Vector Machines (SVMs), for the purpose of emotion recognition within this extensive dataset.

## II. RELATED WORK

this research [1] is to investigate the utility of virtual reality (VR) in the assessment and training of emotion recognition within social contexts, utilizing realistic and dynamic stimuli and involves a comparative analysis of three emotion recognition tasks: VR, video, and photo tasks. In [1] 100 healthy participants who completed all three emotion recognition tasks and during the VR task, participants evaluate emotions expressed by virtual characters (avatars) in a VR urban setting, while their eye movements are tracked. The overall recognition accuracy stands at 75%, aligning closely with the accuracy observed in the photo and video tasks. Nevertheless, distinctions emerge in recognizing specific emotions; VR performs less effectively in identifying disgust and happiness but outperforms the video task in recognizing surprise and anger. Participants allocate varying amounts of time to different emotions during the VR task. Notably, disgust, fear, and sadness receive more attention compared to surprise and happiness. Additionally, participants exhibit a preference for focusing on the eyes and nose regions of the avatars rather than their mouths [1]. In [2] the author acknowledges the distinctive nature of data collection within virtual reality (VR) environments, particularly through head-mounted displays, when contrasted with traditional classroom or online learning settings. This underscores the necessity for developing a recognition approach tailored specifically to VR contexts and they have a novel method for recognizing learning concentration within VR environments. This innovative approach hinges on the integration of multi-modal features, incorporating data derived from learner interactions (e.g., interactive assessments, text interactions, clickstream data) and visual cues (e.g., pupil facial expressions and eye gaze). These combined features facilitate the evaluation of learners' concentration from cognitive, emotional, and behavioral perspectives and the study of this research reveals a positive association between heightened levels of concentration and enhanced learning achievements. Additionally, it highlights the significant role played by learners' perceived sense of immersion within the VR environment in shaping their level of concentration. In [3] the paper describes the creation of an inventive VR-based system designed for presenting facial emotional expressions. This system not only enables the monitoring of emotional expressions but also tracks eye gaze and physiological signals connected to the identification of emotions, introducing more efficient therapeutic methodologies and an evaluation of the new VR-based system was carried out through a usability study involving ten adolescents with ASD and ten typically developing adolescents as a control group. The study involved an analysis of eye tracking data and physiological responses to discern variations in gaze and physiological patterns within and between the two groups. this research [3] indicates that adolescents with ASD have distinct ways of processing and recognizing emotional faces when compared to their typically developing peers. These distinctions are reflected in performance metrics, eye tracking parameters, and physiological markers.

## III. PROPOSED METHOD

In this section, we will discuss the proposed methodology of facial emotion recognition using the facial points gathered from VR device. The below figure is the proposed architecture

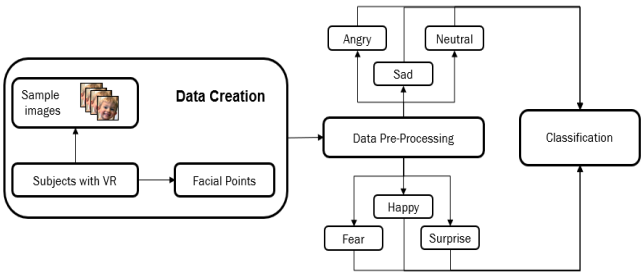



Figure 1 The Proposed Flow Diagram

### *A. Data Pre-processing and Standardization*

The raw data, while invaluable, often demands careful preprocessing to ensure uniformity and compatibility for subsequent analyses. To this end, our preprocessing pipeline encompassed several crucial steps:

Data Tagging: We meticulously annotated the collected data, associating each instance with the respective emotion category elicited by the sample images. This crucial step imbued our dataset with the essential ground truth labels necessary for supervised machine learning.

Dimensional Consistency: Given the variations in data collection, we applied rigorous dimensionality reduction techniques to standardize the data across all CSV files. This process ensured that each file adhered to a uniform structure, facilitating seamless integration and analysis.

### *B. Machine Learning (ML)*

With the standardized and categorized dataset in hand, we proceeded to employ various machine learning algorithms, with a predominant focus on Support Vector Machines (SVM). SVMs, renowned for their versatility and proficiency in classification tasks, were instrumental in the context of emotion recognition. These algorithms were trained on the pre-processed data, leveraging the extracted facial features to make predictions regarding the emotional states expressed by the subjects.

## IV. Results and Discussion

### *A. Dataset*

To obtain a comprehensive understanding of human emotions in immersive contexts, we leveraged VR technology as a potent tool for data collection. Our study involved eight subjects who actively participated in emotion elicitation sessions within the VR environment. Each subject was presented with a set of sample images, meticulously designed to evoke six distinct emotional states: happiness, sadness, anger, neutrality, surprise, and fear. The VR device captured their facial expressions and responses as they viewed and reacted to these visual stimuli.

The resultant data, reflective of real-time emotional responses, was recorded and saved in the form of structured Comma-Separated Values (CSV) files. Each CSV file encapsulated a wealth of information, comprising more than 2000 rows and 64 columns. These columns were thoughtfully named to encompass a broad spectrum of facial movements and expressions, thus providing a nuanced perspective of the subjects' emotional states. Notable columns include indicators of eyebrow movements, cheek puffing, lip actions, jaw dynamics, and gaze directions, among others, making up a comprehensive repertoire of facial cues.

### *B. Experimental results*

In our experimental trials, we attained remarkable levels of accuracy when employing our machine-learning model on the dataset obtained from the VR device. This substantial accuracy underscores the model's proficiency in effectively discerning and categorizing emotions using facial key points as features. These outcomes portend favorable opportunities for the utilization of this approach in diverse applications, notably within the domain of human-computer interaction.

Table 1 above provides a comprehensive summary of the performance metrics for three distinct classifiers employed in our study: cubic SVM, quadratic SVM, and medium Gaussian SVM.

Table I Results of the Different Classifiers

| Classifiers | Recall Rate (%) | Precision Rate (%) | F1 Score (%) | Time (sec) | Accuracy (%) |
|---|---|---|---|---|---|
| **Cubic SVM** | 99.1 % | 99.08 % | 99.09 % | 828 Sec | **99.1 %** |
| **Quadratic SVM** | 97.58 % | 97.63 % | 97.60 % | 630.05 Sec | 97.6 % |
| **Medium Gaussian SVM** | 95.8 % | 95.8 % | 95.95 % | 786.34 Sec | 95.8 % |

These models were rigorously evaluated across various metrics, including recall rate, precision rate, F1 score, processing time, and accuracy, to assess their efficacy in the context of emotion recognition based on facial key points. Notably, the cubic SVM classifier emerged as the top-performing model, achieving the highest accuracy among the classifiers tested.

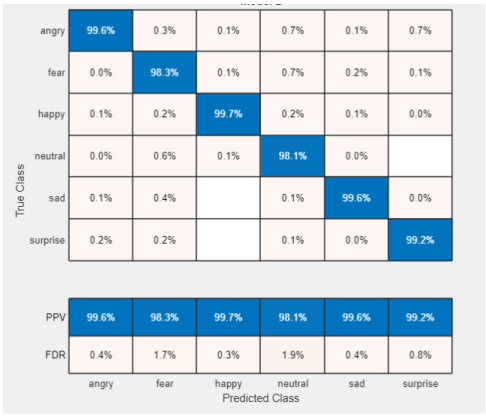



Figure 2 Cubic SVM Recall Rate

Furthermore, it is imperative to note that the success of the cubic SVM classifier extends beyond accuracy, as it consistently demonstrated superior performance in multiple metrics, including recall rate and F1 score, signifying its ability to effectively identify and classify emotions. The classifier's efficiency in processing time also underscores its real-world applicability for interactive systems. For a more detailed understanding of the classification performance, we provide the confusion matrix for the recall rate in Figure 2. And for the precision rate we have provided it in the Figure 3 below.

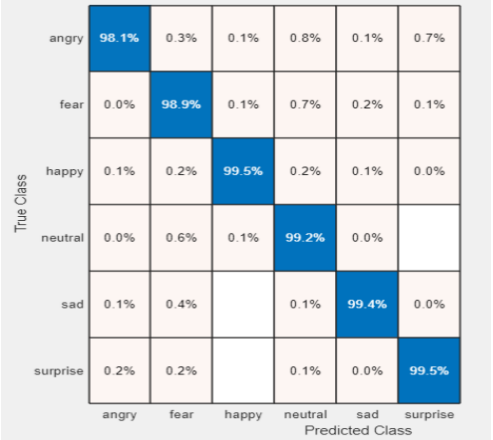



Figure 3 Cubic SVM Precision Rate

The confusion matrix offers a comprehensive view of the classifier's true positive, true negative, false positive, and

false negative predictions, allowing for a deeper insight into its precision and recall rates. These results collectively emphasize the cubic SVM classifier's suitability for applications necessitating accurate and efficient emotion recognition.

## V. CONCLUSION

In this paper, Facial emotion recognition plays an essential role in human-computer interaction, affective computing, and various other domains. In this study, we propose a state-of-the-art approach to facial emotion recognition we present a novel approach to facial emotion recognition the initial phase of our research involved the acquisition of emotional data utilizing a Virtual Reality (VR) device. We enlisted eight participants to contribute to this data collection process, tasking them with displaying predetermined emotions while exposed to a set of sample images. These sample images encompassed six distinct emotional classes, including happiness, sadness, anger, neutrality, surprise, and fear.

The VR device was instrumental in capturing and storing the data in CSV files, resulting in datasets exceeding 2000 rows and comprising 64 columns. Each column was meticulously categorized to correspond to specific facial movement and expression metrics. Subsequently, we diligently undertook data tagging and pre-processing, ensuring uniform dimensions across all files, a critical step in facilitating consistent and reliable analyses. our machine learning model, particularly the cubic Support Vector Machine (SVM), demonstrated exceptional performance. We achieved a remarkable accuracy rate of 99.1% when applying this model to the VR dataset for emotion recognition based on facial key points. This extraordinary level of accuracy underscores the robustness and efficacy of our approach. Such outstanding results hold great promise for the application of our methodology across a broad spectrum of fields, with particular relevance to the advancement of human-computer interaction. Our research then proceeded to employ diverse machine learning algorithms, with a primary focus on Support Vector Machines (SVMs), for the purpose of emotion recognition within this extensive dataset.

## ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ICAN(ICT Challenge and Advanced Network of HRD) program(IITP-2023-2020-0-01832) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation)

## REFERENCE

[1] Geraets, C. N. W., S. Klein Tuente, B. P. Lestestuiver, M. Van Beilen, S. A. Nijman, J. B. C. Marsman, and W. Veling. "Virtual reality facial emotion recognition in social environments: An eye-tracking study." *Internet interventions* 25 (2021): 100432.

[2] Hu, Renhe, Zihan Hui, Yifan Li, and Jueqi Guan. "Research on Learning Concentration Recognition with Multi-Modal Features in Virtual Reality Environments." *Sustainability* 15, no. 15 (2023): 11606.

[3] Bekele, Esubalew, Zhi Zheng, Amy Swanson, Julie Crittendon, Zachary Warren, and Nilanjan Sarkar. "Understanding how adolescents with autism respond to facial expressions in virtual reality environments." *IEEE transactions on visualization and computer graphics* 19, no. 4 (2013): 711-720.

[4] Gutiérrez-Maldonado, José, Mar Rus-Calafell, and Joan González-Conde. "Creation of a new set of dynamic virtual reality faces for the assessment and training of facial emotion recognition ability." *Virtual Reality* 18 (2014): 61-71.

[5] Bisogni, Carmen, Aniello Castiglione, Sanoar Hossain, Fabio Narducci, and Saiyed Umer. "Impact of deep learning approaches on facial expression recognition in healthcare industries." IEEE Transactions on Industrial Informatics 18, no. 8 (2022): 5619-5627.

[6] Tang, Yichuan. "Deep learning using linear support vector machines." arXiv preprint arXiv:1306.0239 (2013).

[7] Goodfellow, Ian J., Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski et al. "Challenges in representation learning: A report on three machine

[8] Bekele, Esubalew, Zhi Zheng, Amy Swanson, Julie Crittendon, Zachary Warren, and Nilanjan Sarkar. "Understanding how adolescents with autism respond to facial expressions in virtual reality environments." IEEE transactions on visualization and computer graphics 19, no. 4 (2013): 711-720.

[9] Sun, Yi, Yuheng Chen, Xiaogang Wang, and Xiaoou Tang. "Deep learning face representation by joint identification-verification." Advances in neural information processing systems 27 (2014).

[10] Ruan, Delian, Yan Yan, Shenqi Lai, Zhenhua Chai, Chunhua

[11] Shen, and Hanzi Wang. "Feature decomposition and reconstruction learning for effective facial expression recognition." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 7660-7669. 2021.

[12] Hua, Wentao, Fei Dai, Liya Huang, Jian Xiong, and Guan Gui. "HERO: Human emotions recognition for realizing intelligent Internet of Things." IEEE Access 7 (2019): 24321-24332.

[13] Zhu, Qing, Qirong Mao, Hongjie Jia, Ocquaye Elias Nii Noi, and Juanjuan Tu. "Convolutional relation network for facial expression recognition in the wild with few-shot learning." Expert Systems with Applications 189 (2022): 116046.

[14] Miranda, Catarina Runa, and Verónica Costa Orvalho. "Assessing Facial Expressions in Virtual Reality Environments." In VISIGRAPP (3: VISAPP), pp. 488-499. 2016.

[15] Zheng, Lim Jia, James Mountstephens, and Jason Teo. "Four-class emotion classification in virtual reality using pupillometry." Journal of Big Data 7 (2020): 1-9.

[16] Dalili, M.N., Penton-Voak, I.S., Harmer, C.J., Munafo, ` M.R., 2015. Meta-analysis of emotion recognition deficits in major depressive disorder. Psychol. Med. 45 (6), 1135–1144

# Novelty of different distance approach for multi-criteria decision-making challenges using q-rung vague sets

*Murugan Palanikumar[1], Nasreen Kausar[2], [*]Dragan Pamucar[3,4], Seifedine Kadry[5,6,7], Chomyong Kim[8], Yunyoung Nam[8],*

[1]*Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai-602105, India.*

[2]*Department of Mathematics,Faculty of Arts and Science, Yildiz Technical University, Esenler, 34220, Istanbul, Turkey.*

[3]*Department of Operations Research and Statistics, Faculty of Organizational Sciences, University of Belgrade, 11000, Belgrade,Serbia.*

[4]*College of Engineering, Yuan Ze University, Taiwan*

[5]*Department of Applied Data Science, Noroff University College, Kristiansand, Norway.*

[6]*Artificial Intelligence Research Center (AIRC), Ajman University, Ajman 346, United Arab Emirates.*

[7]*Department of Electrical and Computer Engineering, Lebanese American University, Byblos 1102-2801, Lebanon.*

[8]*Department of ICT Convergence, Soonchunhyang University, Asan 31538, Korea.*

*Abstract*— **In this article, multiple attribute decision-making problems are solved using the vague normal set (VNS). It is possible to generalize the vague set (VS) and q-rung fuzzy set (FS) into the q-rung vague set (VS). A log q-rung normal vague weighted averaging (log q-rung NVWA), a log q-rung normal vague weighted geometric (log q-rung NVWG), a log generalized q-rung normal vague weighted averaging (log Gq-rung NVWA), and a log generalized q-rung normal vague weighted geometric (log Gq-rung NVWG) operator are discussed in this article. A description is provided of the scoring function, accuracy function and operational laws of the log qrung VS. The algorithms underlying these functions are also described. A numerical example is provided to extend the Euclidean distance and the Humming distance. Additionally, idempotency, boundedness, commutativity, and monotonicity of the log q-rung VS are examined as they facilitate recognizing the optimal alternative more quickly and help clarify conceptualization. We chose five anemia patients with four types of symptoms including seizures, emotional shock or hysteria, brain cause, and high fever, who had either retrograde amnesia, anterograde amnesia, transient global amnesia, post-traumatic amnesia, or infantile amnesia.Natural numbers q are used to express the results of the models. To demonstrate the effectiveness and accuracy of the models we are investigating, we compare several existing models with those that have been developed.**

**Keywords:** Vague set, aggregating operators, Euclidean distance, Hamming distance, decision making

## I. INTRODUCTION

Decision-makers find it increasingly difficult to identify the optimal solution as real-world systems become increasingly complex. Selecting the best option is possible despite the difficulty of deciding between the alternatives. Opportunities, objectives, and viewpoint constraints are challenging to create for many firms. In line with this, when decisions making (DM), individuals or groups should consider multiple objectives at the same time. A wide variety of MADM-related issues are dealt with every day. Our DM abilities need to be improved as a result. This field of study has been studied by a variety of researchers using a variety of methods. There are several uncertain theories proposed by them to deal with the uncertainties,including fuzzy set (FS) [1], intuitionistic fuzzy set (IFS) [2], interval valued FS (IVFS) [3], vague set [4],Pythagorean fuzzy set (PFS) [5], IVPFS [6], spherical FS (SFS) [7]. A membership grade (MG) indicates how well a FS fits into the specified set with ranging from 0 to one. An IFS is defined by Atanassov [2] as having a total of membership grade (MG) and non-membership grade (NMG) less than one. The sum of the MG and NMG is sometimes greater than one when a DM method is applied. Yager of [5] developed PFS, which is characterized by a square sum of its MG and NMG not exceeding one. In order to generalize IFS,Yager used PFS to build a model. A new concept has been proposed by Yager [8] in light of society ' s continuous complexity and theory development. The MG and NMG in the q-rung orthogonal pair FS (q-ROFS) have power q, but the sum can never exceed one. The IFSs and PFSs can all be considered special cases of q-ROFSs, therefore they are general. There are more orthopairs that meet the bounding constraint as rung q increases, and as rung q increases, the space of acceptable orthopairs increases. The use of q-ROFSs can thus express fuzzy information in a broader range. Because the parameter q can be adjusted, q-ROFSs are flexible and better suited to uncertain environments. An increase in q can be made as ambiguity in decision information increases. It is possible that some experts are influenced by both their own desires and their surroundings. Therefore, they may have an MDG of 0.95 and an NMG of 0.55 when evaluating certain decision-making things. The fuzzy information cannot be described by IFNs and PFNs, but q-ROFNs can be described if parameter q is increased. Due to this, the q-ROFS is more flexible and suitable for describing uncertain data.

A discussion of the q-Rung Orthopair fuzzy weighted Archimedean Bonferroni mean (q-ROFWABM) and q-Rung Orthopair fuzzy Archimedean Bonferroni mean (q-ROFABM) operators is given in Liu et al. [9].Liu et al. [10] proposed a concept of q-rung orthopair fuzzy power averaging (q-ROFPA), q-rung orthopair fuzzy power weighted average (q-ROFPWA), q-ROFPMSM and q-rung orthopair fuzzy power

weighted MSM (q-ROFPWMSM) operators for q-ROFNs, describing their properties. Liu et al. [11] discussed the q-rung orthopair fuzzy weighted average (q-ROFWA) and q-rung orthopair fuzzy weighted geometric (qROFWG) operators are introduced and their basic properties are discussed. The concept of an incomplete probabilistic linguistic preference relation (InPLPR) was introduced by Wang et al. [12]. In 2013, Liu et al.[13] presented the concept of unit cost consensus adjustment based on a group consensus decision model based on InPLPR that takes into account social trust networks, consistency, and social trust networks. In a recent study, Zhang et al. [14] discussed three types of multi-granularity q-rung orthopair fuzzy preference relations (PRS) as well as their interesting properties. With the MAGDM algorithm based on q-ROF multiattribute rules, the MG-3WD approach can also be applied to q-ROF complex information systems. Zhang et al. [15] analyzed a UCI dataset using MGq-ROF PRSs, the MULTIMOORA method, and the TPOP method using the MAGDM method. Zhang et al. [16] discuss neutosophic fusion of RST based on basic models and soft sets models. Based on fuzzy granularity spaces with properties that correspond to fuzzy knowledge distances, Lian et al. [17] discuss fuzzy relative knowledge distances. Furthermore, it has been demonstrated that fuzzy knowledge distances contain different structure information than precise knowledge distances.The hybridization of archimedean copulas and generalized MSM operators, based on q-rung probabilistic dual hesitant fuzzy sets, was discussed by Anusha et al. [18]. Multi-attribute decision-making (MADM) [19, 20] offers an efficient means of evaluating multiple alternatives based on their evaluation values. Usually, MADM problems can be solved in one of two ways. Traditional approaches, such as TOPSIS, VIKOR, ELECTRE, are examples. Information integration problems are more effectively solved by AOs than by traditional approaches. In contrast to traditional approaches, AOs provid comprehensive values of all alternatives, rather than ranking results only. In his article, Bairagi [21] used extended TOPSIS to select homogeneous groups of robotic systems.

This is insufficient for demonstrating neutrality (neither favor nor disfavor). It was developed by Cuong et al. [22] with a total grade no higher than one for three pointers such as positive, neutral, and negative. As a result, it would be appropriate for the DM method to use this set over IFS or PFS for selective applications. Liu et al. first presented the concept of an aggregation operator (AO) in generalized PFS

[23]. A PIVFS algorithm for the problem of identifying truth membership grades (TMGs), indeterminacy membership grades (IMGs) and false membership grades (FMGs) with AOs [6, 24, 25, 26] have the feature that the sum of the three grades (TMG, IMG, and FMG) is greater than one. It has been suggested by Ashraf et al. [7] that the SFS should contain the following graph: this diagram shows that the sum of the squares of the TMG, IMG and FMG should be not exceeds one. To analyze the idea of SFS, Fatmaa et al. [27] used the TOPSIS technique as part of their study. There have been several different concepts of q-Rung picture FS with AO for DM that have been demonstrated by Liu et al. [28]. In addition to Gau et al. [4], there is a concept of VSs developed by Gau et al. VS is called to the two functions TMG Tv and FMG Fv as well as a set of transformations. Suppose that Tv(x) is the total likelihood estimate of x, derived from the evidence for x and Fv(x) is the total likelihood estimate for x derived from the evidence against x. It can be noted that these functions fall into the interval [0,1], where their sum is less than one. Various extensions have been made to the VS such as the IVFS and the FS [29, 30, 31]. Zhang et al. [32] was first introduced that suggested PFS can be extended to multi-criteria decision making (MCDM) using TOPSIS. The application of the bipolar fuzzy soft set (BFSS) was explored by Jana et al. [33] for the purpose of discovering how to broaden the set of bipolar fuzzy terms. Ullah et al. [34] describes how pattern recognition applications can be used to estimate PFS distances using complex separation algorithms. It has been discussed that MCDM can be utilized using the neutrosophic set as well as the Dombi power AOs [35]. A number of algebraic structures and their applications have been investigated by Palanikumar et al. [36, 37]. The notion of fuzzy c-number clustering procedures for fuzzy data was discussed by Yang et al. [38, 39].

As an alternative to algebraic operations, a log q-rung arithmetic operation can provide a smooth estimate quality that is similar to that of a continuous algebraic operation when compared with its smoothness. Compared to the log q-rung arithmetic operations on the IFS and PFS only a limited research has been done on log q-rung arithmetic operations. Our method of VS is based on log q-rung arithmetic AOs within VSs rather than using log q-rung arithmetic operations. The use of spherical fuzzy q-rung arithmetic AOs based on entropy in DM, as well as their real-life application to the problem, were introduced by Yun et al.[40]. Ashraf [41] proposed by that linear-logarithmic hybrid AOs be used for single-valued neutrosophic sets. Palanikumar et al. have examined the new type Pythagorean fuzzy set with AO [42]. Yager [5] has also presented an average and geometric AO using PFS weighted and weighted power cases. A number of basic PFS features are discussed by Peng et al. [43]. A generalized PFS under AO has been developed by Liu et al. [23]. Adak et al. dicussed the concept of spherical distance measurement method for solving MCDM problems under PFS [44]. Some picture fuzzy mean operators and their applications in DM is discussed by Hasan et al. [45]. Mishra et al. [46] discussed the new concept of Pythagorean and Fermatean fuzzy sub-group redefined in context of T-norm and S-conorm. DM analysis of minimizing the death rate due to covid-19 by using q-rung orthopair fuzzy soft bonferroni mean operator discussed by Abbas et al. [47]. Yaman [48] discussed the new approach for warehouse location decisions changed in medical sector after pandemic study. Recently, FS and its extension including q-rung orthopair fuzzy set, T-spherical fuzzy set based on decision making approach [49, 50, 51, 52, 53, 54, 55, 56]. The log q-rung information about the VNS was obtained utilizing OAs. This section 2 explains the given information about the FS and VS components. Section 3 explains the definition of q-rung vague sets as well as the different operations involved with them. There is a discussion on ED and HD in section 4 using the log q-rung vague normal number (log q-rung VNN). A MADM connection is established through the section 5 using log q-rung VNNs. Section 6 contains a numerical example and a description of log q-rung

VS as well as the insert algorithm and log q-rung VS application. We provide a conclusion in section 7. An overview of the key things that were taken into account during the research process is given below.

1. As a result of log-rung VNSs, ED and HD were introduced.

2. The log q-rung VNVWA, log q-rung VNVWG, log Gq-rung VNVWA, and log Gq-rung VNVWG operators were suggestions.

3. A log-rung VNS is used in order to explore the MADM technique.

4. We evaluate log q-rung VNVWA, log q-rung VNVWG, log Gq-rung VNVWA and log Gq-rung VNVWG in order to establish optimal value parameters.

5. An analysis of the proposed and early investigations is presented along with a comparative analysis.

6. DM outcomes for natural numbers with a value of q.

## II. Related Work

*Selection of amnesia patients*

A person with anemia has insufficient or malfunctioning red blood cells. A man is diagnosed with anemia when his hemoglobin value is below 13.5 gm/dl, while a woman is diagnosed with anemia when her hemoglobin value is below 12.0 gm/dl. There are a variety of normal values for children depending on their age. Memory strategies are used to help deal with amnesia. Taking care of underlying diseases that cause amnesia is also important. An occupational therapist may help the person learn new information and replace what they have lost. Taking in new information may be based on intact memories. Additionally, memory training can help organize information to make it easier to remember and to better understand when you are speaking with others. Smartphones and hand held tablets are often used by people with amnesia. A simple electronic organizer can help even people who suffer from severe amnesia stay on top of their daily activities with a little training and practice. A person with amnesia may benefit from psychological therapy or cognitive behavioral therapy (CBT). When it comes to recalling forgotten memories, hypnosis can be very effective. It is important to retrieve memories and deal with psychological issues that may have contributed to amnesia as part of treatment for amnesia. A person may be able to retrieve forgotten memories through meditation and related mindfulness activities. It is also imperative to have the support of your family. Playing familiar music, showing them photographs from the past, and exposing them to familiar scents may be helpful. Blood buildup in the brain may cause amnesia in people who have been injured in head trauma. Anti-inflammatory medications may be needed by people with encephalitis. If you cycle, skate, ski, or play contact sports, you may be at greater risk of developing amnesia due to headgear. It is important to consume a diet rich in leafy green vegetables and avoid saturated fats to prevent cardiovascular diseases.that can negatively affect memory. Brain regulation is achieved through the Bilateral Sounds method. As well as relieving stress and symptoms of PCS and PTSD, it is excellent for reeducating the left and right hemispheres. There are some commonly used bilateral sounds available for this purpose through Psych Innovations, a web-based company.

*1. Retrograde amnesia (A):*

A person suffering from retrograde amnesia is incapable of recalling past events.Memory loss usually affects memories made recently, not ones from years ago. You can experience amnesia if you lose the ability to make, store, and retrieve memories. Memory formation prior to amnesia onset is affected by retrograde amnesia. After a traumatic brain injury, a person may develop retrograde amnesia, which prevents him or her from remembering what happened decades earlier. A variety of brain regions can be damaged, causing retrograde amnesia to occur.

*2. Anterograde amnesia (B):*

The type of amnesia that causes this is when you forget anything that has happened since your amnesia began. Even if you have a state of amnesia, you can still recall information you recall before the amnesia occurred. Unlike retrograde amnesia, this occurs more frequently. During an amnesiainducing event, there is no memory creation after anterograde amnesia occurs. It is possible to suffer from anterograde amnesia either to the extent of being unable to remember events only partially or completely. In this case, a person with amnesia has retained long-term memories from the time before the incident occurred. In anterograde amnesia, new memories cannot be encoded (or possibly retrieved). As well as different severity levels of anterograde amnesia, some individuals forget recent events such as meals or phone numbers, while others forget what they were doing a few seconds ago. Memory is also affected by the difficulty of a task, with more complex tasks being harder to remember than simpler tasks that do not require as much mental energy.

*3. Transient global amnesia (TGA) (C):*

It tends to resolve within 24 hours if it is a temporary amnesia. Adults over the age of middle age and those who are older are more likely to experience it. It is rare for such amnesia to recur once it has resolved. Someone who is otherwise alert may experience transient global amnesia, which manifests itself suddenly as confusion. A person with transient global amnesia cannot create new memories, so the memory of recent events is lost. This condition is not caused by something more common, such as epilepsy or stroke. Neither you nor how you got here can be recalled. What’s going on right now may not be clear to you. The answers you’ve just been given may not stick in your memory, so you keep repeating the same questions. Similarly, it is possible to lose track of events from a month ago if you are asked to recall them. People in their middle and older years are most likely to suffer from this condition. Transient global amnesia also leaves you recognizing people you know and remembering who you are. There are always a few hours of recovery time after an episode of transient global amnesia. Your memory may begin to return during recovery. It’s not dangerous, but transient global amnesia can still be frightening.

*4. Post-traumatic (D):*

Amnesia can occur either anterogradely or retrogradely

after an injury to the head. Post-traumatic amnesia is a type of memory loss that occurs immediately after a traumatic brain injury (TBI). This state is characterized by disorientation and inability to remember past events. An individual may be incapable of stating their name, location, and time. It is considered that PTA has been resolved when continuous memory returns. The memory is not able to store new events during PTA. The memory of some incidents is only recalled by one third of patients with mild head injuries. There is a "clouding" of consciousness experienced by the patient during PTA. It has been proposed as an alternative term for PTA since it includes confusion along with the memory loss typically associated with amnesia.

*5. Infantile amnesia (E):*

Children often have difficulty recalling early childhood memories, which is referred to as childhood amnesia. The brains of young children are still developing, so they are incapable of consolidating memories. The ailment of being unable to recall episodic memories in adults younger than two to four years of age is known as childhood amnesia. During these years, the recollection of early childhood memories may also be scarce or fragmented, especially if they occurred between the ages of 2 and 6. Others believe that early memories are encoded and stored differently when a cognitive self is developed. The onset of childhood amnesia has differed between psychologists, but some research shows that children can recall things before they are two years old. As children grow, their memories may decline. An individual can recall their first memory at a certain age, according to some definitions. As a general rule, it occurs at the age of two to four, but this can vary from child to child.

The four factors are

*1. Seizures (e1):*

The underlying mechanisms of seizures are poorly understood, which leads to retrograde amnesia. It was determined whether seizures activate neurons that overlap with engrams of spatial memory and if seizures saturate LTP in engram cells. Retrograde amnesia was caused by a seizure for spatial memory tasks. Bilateral mesiotemporal lesions in humans can cause anterograde amnesia, a severely disabling state. An episode of retrograde and / or anterograde amnesia is characteristic of transient epileptic amnesia (TEA). In the event of a traumatic brain injury (TBI), posttraumatic amnesia may result in confusion and memory loss. This period can be characterized by seizures, but they are not common during this period. In general, seizures are more common after a TBI, during the acute phase immediately after the injury. This stage of the brain's development is when significant changes are occurring, and seizure activity is more likely to occur. Depending on the severity of the injury,

seizures can also occur during the PTA phase. Infantile amnesia is currently not believed to be caused or contributed to by seizures during the period of infantile amnesia. The consequences of seizures on memory and cognition can be severe, especially if they occur at critical periods in a person's life. The consequences of repeated seizures on the brain include changes in the structure and function of the brain that can lead to mental impairments in the long run.

*2. Emotional shock or hysteria (e2):*

Patients who cannot recall particular past events or those that occur during a particular period of their lives suffer from one type of memory loss. There appears to be no connection between retrograde amnesia and any particular brain disorder, past or present. A triggering event for anterograde amnesia can be emotional shock or hysteria, which prevent the brain from processing or retaining memories of the event. A 'wandering womb' is thought to be responsible for hysteria, which is a set of symptoms common to women. Medical practitioners no longer use it as a diagnosis due to its discreditation. There is no connection between hysteria and PTA, which is caused by physical trauma to the brain. Hysteria and infantile amnesia do not have any direct connection. The terms refer to different phenomena within the brain, though they both involve the functioning of the brain. Hysteria is no longer used as a diagnosis because infantile amnesia is a normal development stage. Discussing mental health and neurological conditions requires accurate and current terminology.

*3. Brain Cause (e3):*

Memory-storing areas of the brain in several brain regions can be damaged, resulting in retrograde amnesia. There are numerous different causes of this type of damage, including trauma, serious illnesses, seizures, strokes, and degenerative brain diseases. Alzheimer's disease and frontotemporal dementia are the two conditions that cause anterograde amnesia most often. When your brain deteriorates and stops functioning, memory loss is extremely common. An amnesia that lasts for several hours is known as transient global amnesia (TGA). A temporary disruption of blood flow and oxygen to certain parts of the brain, particularly the hippocampus, is thought to be the cause of TGA, but its exact cause is unknown. Located deep within the brain, the hippocampus is a small seahorse-shaped structure that plays a crucial role in forming new memories as well as retrieving old ones. There are several factors that can interfere with the flow of blood and oxygen to this area. Post-traumatic amnesia (PTA) can occur following trauma to the brain (TBI). When the brain is injured, it can disrupt the normal functioning of the brain cells, known as neurons. Damage to the brain can occur as a result of a direct impact on the head or from shaking the skull. During the processing of information within the brain, neurons are responsible for transmitting information, and damage to them causes the brain to lose function. Some people believe infantile amnesia is caused by the underdevelopment of the infant brain, which would make consolidation of memory impossible, or by memory retrieval deficits.

*4. High fever (e4):*

Memory loss and confusion are common symptoms of a high fever. Symptoms of high fever can include retrograde amnesia, which is the loss of memory of events that occurred before the fever began. High fevers, head trauma, strokes, and other medical conditions can cause retrograde amnesia, which is a condition caused by damage to the brain. As a result of an incident or injury, anterograde amnesia can occur. Confusion, delirium, and memory problems can all be caused by high fever. In contrast, there is no association between anterograde amnesia and this condition. A person experiencing anterograde

amnesia is most likely suffering from damage to the brain regions involved in forming and consolidating new memories. Memory problems can occur as a result of high fevers, such as those caused by encephalitis or meningitis. TGA is not known to cause fever, but high fevers are sometimes experienced by people with TGA. It can take from minutes to weeks or even months to recover from a PTA injury, depending on the severity. An individual with a TBI may experience a fever during the acute phase. The body can fight off pathogens and promote healing when it experiences fever due to infection or injury. In contrast, a high fever or prolonged fever can cause further damage to the brain and other complications if it is very high or prolonged. It is important to monitor infants and young children closely when they have a fever. Medical attention should be sought if the fever has an underlying cause. Infections caused by viruses or bacteria, teething, and immunizations all contribute to fever in infants. Medications such as acetaminophen and ibuprofen may be given to reduce fever, and hydration may be encouraged. Suppose that five anemia patients as $\mathscr{P}$ = {A,B,C,D,E}. Four factors are considered as $\gamma$ = {e1, e2, e3, e4} and their weights are $\Psi$ = {0.4,0.3,0.2,0.1}. First aid treatments should be selected for each alternative.

## III. Conclusion

Clearly, this method is effective due to its ability to consider relationships between attributes. As a result, the proposed method produces more accurate ranking results. In consideration of the interrelationships between attributes, the proposed method is more efficient and superior to [26] in solving practical DM problems. In this article, we examined problems arising within DM domains using log q-rung NVS using MADM. Based on our discussion of log q-rung NVS, several AO reached a number of conclusions that were important to their log q-rung NVS. There should be a log q-rung NVWA and log q-rung NVWG, as well as a log G q-rung NVWA and log G q-rung NVWG. By applying log q-rung NVS based on the MADM methodology, individuals may be able to determine the appropriate action to take in scenarios with unclear and contradictory facts. We apply the operator representations of log q-rung NVWAs, log q-rung NVWGs, log G q-rung NVWAs and log G q-rung NVWGs to problems, which are based on log q-rung NVS. We can estimate the different rankings using log q-rung NVWA, log q-rung NVWG, log G q-rung NVWA, and log G q-rung NVWG. As a final step, we have examined the values of q that affect alternative ranking most strongly. A decision-maker can select the most appropriate ranking based on a real-world scenario by adjusting q. Based on the actual values of q, the decision-maker can select a method. Finally, we compared the proposed models to a number of currently in use models in order to demonstrate their applicability and benefits. In data analysis, HD and ED of neutrosophic sets are used in a number of practical applications. If further research shows that these operators are superior to others, such as power mean aggregation operators, Bonferroni mean operators, Heronian mean operators, etc., we may be able to extend new new q-rung complex neutrosophic set to them. The following topics will be discussed in further detail: (1) It is shown that expert sets and soft sets can be compared with log q-rung NVSs.

## References


[1] Zadeh, L. A. (1965). *Fuzzy sets. Information and control,* 8(3), 338 – 353.

[2] Atanassov, K. (1986). *Intuitionistic fuzzy sets. Fuzzy sets and Systems,* 20(1), 87 – 96.

[3] Gorzalczany, M. (1987). A method of inference in approximate reasoning based on interval valued uzzy sets. *Fuzzy Sets and Systems*, 21, 1 – 17.

[4] Biswas, R. (2006). Vague groups. *International journal of Computational Cognition*, 4(2), 20 – 23.

[5] Yager, R. (2014). Pythagorean membership grades in multi criteria decision-making. *IEEE Trans. Fuzzy Systems,* 22, 958 – 965.

[6] Peng, X., Yang, Y. (2015). Fundamental properties of interval valued pythagorean fuzzy aggregation operators. *International Journal of Intelligent Systems*, 31(5), 1 – 44.

[7] Ashraf, S., Abdullah, S., Mahmood, T., Ghani, F., Mahmood, T. (2019). Spherical fuzzy sets and their applications in multi-attribute decision making problems. *Journal of Intelligent and Fuzzy Systems*, 36,2829 – 2841.

[8] Yager, R. R. (2016). Generalized orthopair fuzzy sets. *IEEE Transactions Fuzzy Systems,* 25(5), 1222 – 1230.

[9] Liu, P., Wang, P. (2018). Multiple-attribute decision-making based on archimedean bonferroni operators of q-rung orthopair fuzzy numbers. *IEEE Transactions on Fuzzy Systems*, 27(5), 834 – 848.

[10] Liu, P., Chen, S. M., Wang, P. (2020). Multiple attribute group decision-making based on q-rung orthopair fuzzy power maclaurin symmetric mean operators.*IEEE Transactions on System and Cybernetics System*, 10(50), 3741 – 3756.

[11] Liu, P., Wang, P. (2017). Some q-rung orthopair fuzzy aggregation operators and their applications to multiple-attribute decision making. *International journal of Intelligent Systems*, 1, 1 – 22.

[12] Wang, P., Liu, P., Chiclana, F. (2021). Multi-stage consistency optimization algorithm for decision making with incomplete probabilistic linguistic preference relation. *Information Sciences*, 556, 361 – 388.

[13] Liu, P., Dang, R., Wang, P., Xiaoming, W. (2023). Unit consensus cost-based approach for group decision-making with incomplete probabilistic linguistic preference relations. *Information Sciences*,624, 849 – 880.

[14] Zhang, C., Ding, J., Li, D., Zhan, J. (2021). A novel multi-granularity three-way decision making approach in q-rung orthopair fuzzy information systems. *International Journal of Approximate Reasoning*,138, 161 – 187.

[15] Zhang, C., Bai, W., Li, D., Zhan, J. (2022). Multiple attribute group decision making based on multigranulation probabilistic models, multimoora and tpop in incomplete q-rung orthopair fuzzy information systems. *International Journal of Approximate Reasoning,* 143, 102 – 120.

[16] Zhang, C., Li, D., Xiangping, K., Song, D., Sangaiah, A. K., et al. (2020). Neutrosophic fusion of rough set theory: An overview. *Computers in Industry*, 115, 103 – 117.

[17] Lian, K., Wang, T., Wang, B., Wang, M., Huang, W., et al. (2023). The research on relative knowledge distances and their cognitive features. *International Journal of Cognitive Computing in Engineering,* 4,135 – 148.

[18] Anusha, G., Ramana, P., Sarkar, R. (2023). Hybridizations of archimedean copula and generalized msm operators and their applications in interactive decision-making with q-rung probabilistic dual hesitant fuzzy environment.*Decision Making: Applications in Management and Engineering,* 6(1), 646 – 678.

[19] Liu, P. D., Teng, F. (2019). Probabilistic linguistic todim method for selecting products through online product reviews. *Information Sciences,* 485, 441 – 455.

[20] Pang, Q., Wang, H., Xu, Z. (2016). Probabilistic linguistic linguistic term sets in multi-attribute group decision making. *Information Sciences,* 369, 128 – 143.

[21] Bairagi, B. (2022). A homogeneous group decision making for selection of robotic systems using extended topsis under subjective and objective factors.

*Decision Making: Applications in Management and Engineering*, 5(2), 300 – 315.

[22] Cuong, B., Kreinovich, V. (2013). Picture fuzzy sets a new concept for computational intelligence problems. *Proceedings of 2013 Third World Congress on Information and Communication Technologies*(WICT 2013), IEEE, 1 – 6.

[23] Liu, W., Chang, J., He, X. (2016). Generalized pythagorean fuzzy aggregation operators and applications in decision making. *Control Decision*, 31, 2280 – 2286.

[24] Rahman, K., Abdullah, S., Shakeel, M., Khan, M., Ullah,M.(2017).Interval valued pythagorean fuzzy geometric aggregation operators and their application to group decision-making problem. *Cogent Mathematics*, 4, 1 – 19.

[25] Rahman, K., Ali, A., Abdullah, S., Amin, F. (2018). Approaches to multi attribute group decisionmaking based on induced interval valued pythagorean fuzzy einstein aggregation operator. *New Mathematics and Natural Computation*, 14(3), 343 – 361.

[26] Yang, Z., Chang, J. (2020). Interval-valued pythagorean normal fuzzy information aggregation operators for multiple attribute decision making approach. *IEEE Access*, 8, 51295 – 51314.

[27] Fatmaa, K. G., Cengiza, K. (2019). Spherical fuzzy sets and spherical fuzzy topsis method. *Journal of Intelligent and Fuzzy Systems,* 36(1), 337 – 352.

[28] Liu, P., Shahzadi, G., Akram, M. (2020). Specific types of q-rung picture fuzzy yager aggregation operators for decision-making. *International Journal of Computational Intelligence Systems*, 13(1),1072 – 1091.

[29] Bustince, H., Burillo, P. (1996). Vague sets are intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 79,403 – 405.

[30] Kumar, A., Yadav, S. P., Kumar, S. (2007). Fuzzy system reliability analysis using t based arithmetic operations on lr type interval valued vague sets. *International Journal of Quality and Reliability Management,* 24(8), 846 – 860.

[31] Wang, J., Liu, S. J., Zhang, J., Wang, S. Y. (2006). On the parameterized owa operators for fuzzy mcdm based on vague set theory. *Fuzzy Optimization and Decision Making,* 5, 5 – 20.

[32] Zhang, X., Xu, Z. (2014). Extension of topsis to multiple criteria decision-making with pythagorean fuzzy sets. *International Journal of Intelligent Systems*, 29, 1061 – 1078.

[33] Jana, C., Pal, M. (2018). Application of bipolar intuitionistic fuzzy soft sets in decision-making problem. *International Journal of Fuzzy System Applications*, 7(3), 32 – 55.

[34] Ullah, K., Mahmood, T., Ali, Z., Jan, N. (2019). On some distance measures of complex pythagorean fuzzy sets and their applications in pattern recognition. *Complex and Intelligent Systems*, 1 – 13.

[35] Jana, C., Pal, M. (2021). Multi criteria decision-making process based on some single valued neutrosophic dombi power aggregation operators. *Soft Computing,* 25(7), 5055 – 5072.

[36] Palanikumar, M., Iampan, A. (2022). Spherical fermatean interval valued fuzzy soft set based on multi criteria group decision making. *International Journal of Innovative Computing, Information and Control*, 18(2), 607 – 619.

[37] Palanikumar, M., Iampan, A. (2022). Novel approach to decision making based on type-ii generalized fermatean bipolar fuzzy soft sets. *International Journal of Innovative Computing, Information and Control,* 18(3), 769 – 782.

[38] Yang, M. S., Ko, C. H. (1996). On a class of fuzzy c-numbers clustering procedures for fuzzy data. *Fuzzy Sets and Systems*, 84, 49 – 60.

[39] Xu, R. N., Li, C. L. (2001). Regression prediction for fuzzy time series. *Appl. Math. J. Chinese Univ.*,16, 451 – 461.

[40] Jin, Y., Ashraf, S., Abdullah, S. (2019). *Spherical fuzzy logarithmic aggregation operators based on entropy and their application in decision support systems.* Entropy, 21, 1 – 36.

[41] Ashraf, S., Abdullah, S., Smarandache, F. (2019). *Logarithmic hybrid aggregation operators based on single valued neutrosophic sets and their applications in decision support systems.* Symmetry, 11,364 – 376.

[42] Palanikumar, M., Arulmozhi, K., Jana, C. (2022). Multiple attribute decision-making approach for Pythagorean neutrosophic normal interval-valued aggregation operators. *Comp. Appl. Math*., 41(90), 1 – 27.

[43] Peng, X., Yuan, H. (2016). Fundamental properties of pythagorean fuzzy aggregation operators. *Fundam. Inform.,* 147, 415 – 446.

[44] Adak, A. K., Kumar, G. (2023). Spherical distance measurement method for solving mcdm problems under pythagorean fuzzy environment. *Journal of Fuzzy Extension & Applications,* 4(1), 28 – 39.

[45] Hasan, M. K., Ali, M. Y., Sultana, A., Mitra, N. K. (2022). Some picture fuzzy mean operators and their applications in decision-making. *Journal of Fuzzy Extension & Applications,* 3(4), 349 – 361.

[46] Mishra, V. N., Kumar, T., Sharma, M. K., Rathour, L. (2023). Pythagorean and fermatean fuzzy subgroup redefined in context of t-norm and s-conorm. *Journal of Fuzzy Extension & Applications,* 4(2),125 – 135.

[47] Abbas, M., Asghar, M. W., Guo, Y. (2022). Decision-making analysis of minimizing the death rate due to covid-19 by using q-rung orthopair fuzzy soft bonferroni mean operator. *Journal of Fuzzy Extension & Applications,* 3(3), 231 – 248.

[48] Yaman, T. T., Akkartal, G. R. (2022). How warehouse location decisions changed in medical sector after pandemic? a fuzzy comparative study. *Journal of Fuzzy Extension & Applications,* 3(1), 81 – 95.

[49] Limboo, B., Dutta, P. (2022). A q-rung orthopair basic probability assignment and its application in medical diagnosis. *Decision Making: Applications in Management and Engineering,* 5(1), 290 – 308.

# A Novel Deep Learning-Based Model for Classification of Wheat Gene Expression

Amr Ismail[1], Walid Hamdy[1,2], Aya M. Al-Zoghby[3], Wael A. Awad[3], Ahmed Ismail Ebada[3], Yunyoung Nam[4], Byeong-Gwon Kang[4] and Mohamed Abouhawwash[5,6]

[1]*Faculty of Science, Port Said University, Port Said, Egypt*

[2]*Modern Academy for Computer Science and Management Technology, Cairo 11742, Egypt*

[3] *Faculty of Computers and Artificial intelligence, Damietta University, New Damietta, Egypt*

[4]*Department of ICT Convergence, Soonchunhyang University, South Korea*

[5]*Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt.*

[6]*Department of Computational Mathematics, Science, and Engineering (CMSE), Michigan State University, East Lansing, MI, 48824 USA.*

*Abstract*— **Deep learning (DL) plays a critical role in processing and converting data into knowledge and decisions. DL technologies have been applied in a variety of applications, including image, video, and genome sequence analysis. In deep learning the most widely utilized architecture is Convolutional Neural Networks (CNN) are taught discriminatory traits in a supervised environment. In comparison to other classic neural networks, CNN makes use of a limited number of artificial neurons, therefore it is ideal for the recognition and processing of wheat gene sequences. Wheat is an essential crop of cereals for people around the world. Wheat Genotypes identification has an impact on the possible development of many countries in the agricultural sector. In quantitative genetics prediction of genetic values is a central issue. Wheat is an allohexaploid (AABBDD) with three distinct genomes. The sizes of the wheat genome are quite l large compared to many other kinds and the availability of a diversity of genetic knowledge and normal structure at breeding lines of wheat, Therefore, genome sequence approaches based on techniques of Artificial Intelligence (AI) are necessary. This paper focuses on using the Wheat genome sequence will assist wheat producers in making better use of their genetic resources and managing genetic variation in their breeding program, as well as propose a novel model based on deep learning for offering a fundamental overview of genomic prediction theory and current constraints. In this paper, the hyperparameters of the network are optimized in the CNN to decrease the requirement for manual search and enhance network performance using a new proposed model built on an optimization algorithm and Convolutional Neural Networks (CNN).**

**Keywords:** Gene expression; convolutional neural network; optimization algorithm; genomic prediction; wheat

## I. INTRODUCTION

Cultivated crops must be increased to meet the world's population's food, feed, and fuel demand projected at more than 9 billion by 2050 [1]. One in nine people currently finds themselves living under food insecurity [2]. With limited opportunities to expand farming on existing land, increasing yields could dramatically reduce the number of people at risk of starvation [3]. Given the need to increase crop production by 50 percent by 2050 [4], our current yield levels are inadequate to achieve this target [5]. Therefore, it is necessary and urgent to find ways to boost crop productivity, such as by genetically modifying cultivars and improving agricultural practices [6, 7]. The plant sector is the center of many countries' production. Growing plant typically has special features, such as habits, morphology, and economic value. According to statistics, several plants are registered and named worldwide [8]. We apply genomic prediction techniques in the plant recognition and identification study to make this industry successful. New approaches and techniques in the detection of plant diseases are being employed in the Genomic processing industry. Therefore, in recent years, researchers have become involved in the detection of plant diseases by using genomic processing technology for their importance and effect on farming's future. However, the prediction of the wheat gene is a new problem in machine learning. Through this method, the goal is to achieve a perfect model for wheat gene expression.

## II. RELATED WORK

Deep learning is developing into a strong form of machine learning, which benefits both the outstanding computational resources and the very large available datasets [9]. The need to specifically define which features to use or use for data analysis is bypassed by deep learning. Deep learning then optimizes a robust end-to-end cycle by mapping data samples to outputs compatible with the large identified network training data sets. The CNNs practice this end-to-end mapping for image processing activities, by optimizing several layers of filters. The first filters are interpreted simply as low-level image features (e.g., borders, bright spots, color variations), and the subsequent layer combinations are more and more complex. CNN greatly outperforms all current alternative methods for image analysis where adequate training is given. Results improved from 84.6 percent in 2012 [10] to 96.4 percent in 2015 [11] with benchmark-classification tasks attempting to determine which one thousand different objects are pictures.

Machine Learning technology have be used in a lot of applications in recent years, including image processing. CNN

as indicated in [12] is the most common architecture and is primarily used in deep analysis. The CNN is equipped with discriminatory learning features by supervised means. In contrast to other conventional neural networks, CNN utilizes a few artificial neurons that make it suitable for image detection and processing. On the other hand, for training phases, CNN needs a broad sample number. CNN also has hyperparameters and a wide range of special architectures that are considered expensive and difficult to identify manually such as optimum hyperparameters [13]. We are responsive to the planning, which dramatically impacts CNN efficiency, of certain hyperparameters. Moreover, the hyperparameters for each dataset have to be modified because the over-parameters are different from one dataset to another. The correct values for hyperparameters for a certain dataset are calculated by trial and error since a math format is not given to manually change the hyperparameters. Selecting hyperparameter values requires detailed data that forces non-experts to use a random search or a grid seeking to find the better hyperparameters, which achieve the best performance of CNN. In [14] They used six deep neural networks and machine learning techniques to investigate and exploit the methylation patterns of the Chinese spring bread wheat cultivar in order to identify differentially expressed genes (DEGs) between leaves and roots. Genes with increased terms at leaves were mostly engaged in pigment and photosynthesis production activities, as expected, whereas genes with no difference in expression amidst leaves and roots were mostly implicated in protein processing and diaphragm structures. In [15] They used this study to see how well the DL model worked in the spring wheat breeding programme at Washington State University. They compared and evaluated the execution of two DL techniques, the convolutional neural network (CNN) and the multilayer perceptron (MLP), ridge retraction better linear equitable predictor (rrBLUP), which is a popular GS model. They used the nested association mapping (NAM) for the Spring wheat many seeded from the 2014–2016 growth seasons yielded 650 recombinant inbred lines (RILs). They used cross-validations (CVs), alternative sets, and independent validations of SNP markers, they made predictions for five various quantitative variables using various genetic architectures. Hyperparameters for models of DL were adjusted by decreasing the root average square in the training dataset and employing dropout and regularization to avoid model overfitting.

In [16] they used R-CNN Faster to verify the spike number by using the dataset for high-density wheat 660K SNP array. they achieved an accuracy of 86.7%. They approve that the R-CNN Faster model is faster and has a high accuracy that may be applied to genetic investigations of SN in wheat.

## III. Deep Learning Principles

A standardized DL architecture consists of a mixture of multiple "neurons" layers. In the 50s, with a prominent "perceptron" of Rosenblatt, the idea of a nerve network was proposed, inspired by the activity of the brain [17]. In the past decade, the DL resurgence was focused on the development of powerful algorithms which can be used in complex network parameters containing multiple layers of neurons (e.g. backpropagation) [18] and on the fact that they surpass current algorithms in various automated recognizing functions like picture checking [19]. Deep learning is an area of many specific jargon terms, which means that some of the most crucial terms are defined in Figure 1 to make understanding easier for an inexperienced user.

Fig. 1 Multistage perceptron (MLP) graph displaying the feedback of the simple "Neuron" with n inputs and four hidden layers of single nucleotide polymorphisms (SNPs). The linear combinations' nonlinear transformations (xi, wi, and biases b) all culminate in a single neuron. where $x_i$ represents the neuron's i input, $w_i$ represents a weight connected by input i, and b represents a time-invariant alignment level.

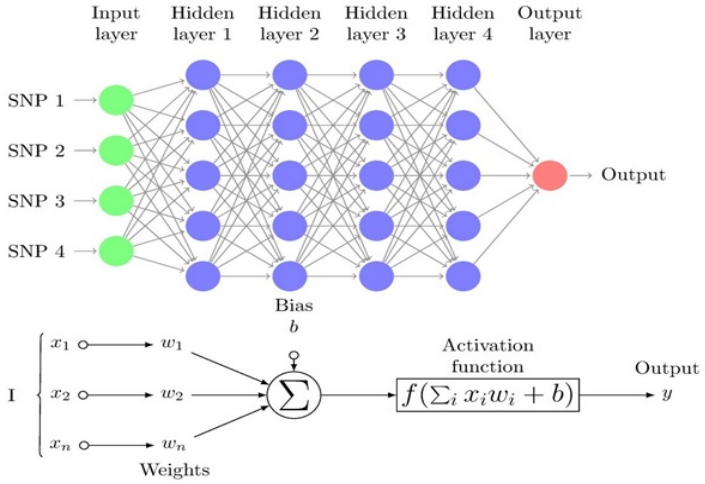



Figure 1: Multistage perceptron (MLP).

The linear combinations' nonlinear transformations (xi, wi, and biases b) all culminate in a single neuron. where $x_i$ represents the neuron's i input, $w_i$ represents a weight connected by input i, and b represents a time-invariant alignment level.

### A. Deep Learning Architectures

Although all DL techniques generally use stacked neuron layers, they do also include a large architecture. The most prevalent ones are convolutional neuro-networks (CNN), multilayer perceptron (MLP), generative opposing networks (GANs), and recurrent neural networks (RNNs). These are listed in effect, although the reader should be aware of various additional options [20].

The multi-layer perceptron network (MLP) consists of a set of completely connected layers named hidden and input layers (see Figure 2) and is one of the most common DL architectures. The first layer receives SNP genotypes (x) feedback in the sense of genomic prediction [21], while the initial layer's output is a weighted, non-linear function of all feedback plus a "bias". Then the first output layer is shown in Eq. (1):

$$z^{(1)} = b_0 + W^{(0)} f^{(0)}(x) \quad (1)$$

When x includes each individual's genotypes, b is considered

a "bias" and is measured along with the remaining weights $W^0$ and f is a nonlinear function (activation function available on Keras). The same term is used in successive layers so that the neuron's inputs of a certain layer are the outputs of the preceding layer $z^{(k-1)}$:

$$z^{(k)}=b_k+W^{(k-1)}f^{(k-1)}(z^{(k-1)}) \quad (2)$$

The final layer generates a number matrix, whether the goal is a true phenotype, or if the goal is a class (ie a problem with classifying) an array of probabilities for each point. Although MLPs constitute a powerful strategy for managing classification or regression issues, they are not the perfect way to handle space or time sets [22]. In latest years, other methods of DL have been suggested in order to deal with these challenges, such as recurrent neural networks, deep generative networks, or convolutional neural networks.

Input variables have been spread in accordance with space models with one dimension (for example, SNPs or text) and two or three dimensions (for example, images), to conform to the circumstances of the implementation of neural networks. CNN's have been introduced. CNN is a particular type of neural network that uses convolution in hidden layers rather than of full matrix reproduction [23]. A CNN consists of thick layers and "convolutional layers" that are fully connected (Fig. 2). An overall operation as well as the input of predetermined width and steps are done in every convolutionary layer. A 'kernel' or 'buffer' is a collection of convolutional processes that functions similarly to a 'neuron' in an MLP [24].

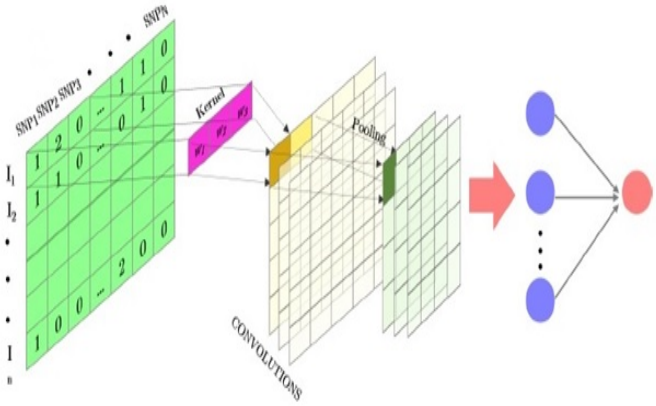



Figure 2: Total view of 1D fully convolutional SNP-matrix neural network.

After each convolution, the output is generated using an activation function. Finally, the results are frequently evened out through a "pooling" method. The kernel outputs of the various positioning positions are combined by using all values of those positions on average, maximum, or minimum. Its capability to which the amount of parameters to be determined is one of the main advantages of convolution networks. These networks have already restricted connections and are translations similar. Fig. 3 provides an example of a one-dimensional (1D) kernel convolution with a scale of 3K [25].

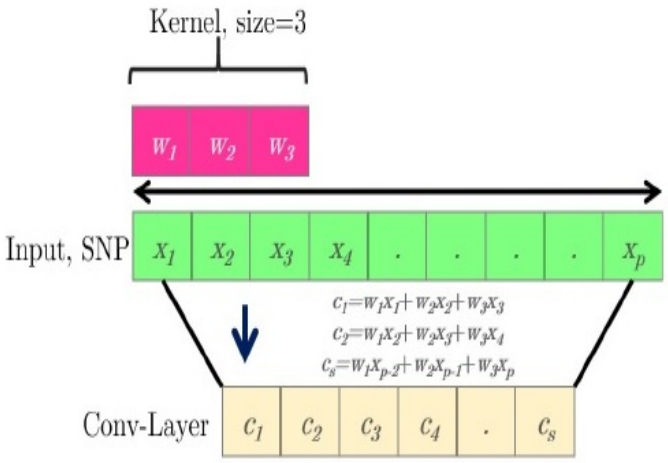



Figure 3: Simple one-dimensional (1D) operation scheme.

### *B. Convolutional Neural Network*

The CNN is so good at categorizing simple patterns in data, it might be utilized to build additional complicated patterns during higher layers. CNNs are a specific type of multilayer neural network. It is trained using the backpropagation algorithm, which is used by practically all other neural networks. CNN's architecture sets it apart from the competition. In a CNN design, there are input layers, numerous hidden levels, and output layers. The hidden layer is made up of pooling layers, Convolutional layers, and fully connected layers [26].

The input data is received by the convolution layer, which applies a filter to it, essentially, the input data is multiplied by the kernel to generate the adjusted output data. A convolution layer subsampling method is the Pooling layer. The goal is to reduce the number of dimensions. An input layer serves as the first layer in the proposed CNN algorithm used in this study. The second layer makes up a one-dimensional convolution layer with three kernel sizes, a 30 filter, and RELU activation. The third layer is the max pooling layer, with two pool sizes. The next layer is a completely connected layer with the ability to activate RELU.

Finally, the output layer is made up of a single sigmoid activation in a neuron. The ADAM optimizer is applied for learning, as a cost function with binary cross-entropy.

### *C. Recurrent Neural Network*

RNN is the only algorithm with internal memory. Therefore, it is a very powerful and reliable algorithm, the RNN is incredibly powerful since it is still the only algorithm with internal memory. The internal memory of the RNN allows the algorithm to recall and research critical information about the input it receives; this ability allows the program to predict what will happen next with great accuracy [27]. The information in an RNN loops back on itself. As demonstrated in Fig. 4, it considers the current input as well as what it has learned from previous inputs before making a decision.

This study employs a simple RNN layer, with the output being fed back into the input. The simple RNN layer is used to apply

the RELU activation function. A sigmoid activation algorithm was also employed for the output layer. For learning, the ADAM optimizer is employed, and as a cost function, binary cross-entropy is used.

## IV. The Proposed Approach

In this section, the dataset used to implement the proposed approach is first described, then the details of the approach proposed are explained.

### *D. Dataset description*

The data for this study is from the Global Wheat data set, which is open to the public [28]. The original genotypic data consisted of 73,345 polymorphic markers anchored to the Chinese Spring Ref Seqv1 map. Before filtering the genotypic data, RILs with lacking phenotypic data in a single setting were deleted. SNP markers having a missing data rate of higher than 20%, minor allele frequencies of less than 0.10, and RILs with more than 10% genotypic data were also eliminated, leaving 40,000 SNP and 635 RILs markers for analysis. Using 635 RILs and 40,000 SNP markers, The demographic structure of the 26 NAM families was investigated using principal component analysis (PCA).

### *E. The Proposed Approach for Classification of Wheat Gene Expression*

The solution suggested is an incredibly effective way of optimizing the efficiency of the CNN network by the incorporation of the terminals of two pre-trained CNN networks. In fact, the model's hyperparameters are designed such that each model performs better.

## V. EXPERIMENTS AND RESULT

The protein interaction network was mapped to the gene expression levels from our dataset. Each time, 95 samples were used as testing data and 285 samples were used as training data to train the convolutional neural network architecture. There were 60 epochs in total. Then, by 95 samples (4x95=380) and k=4, we rank-fold cross-validation (CV). The selecting test data were then randomized to a 10-time process of randomization, after which the average value for the following machine learning metrics—accuracy, specificity, recall (sensitivity), and precision—was calculated.

The size and number of convolutional filters, as well as the number and size of convolutional layers and hidden layers, were all examined in various combinations. With the architecture, the best outcomes were obtained. Tab. 1 shows that with our sample, with a mean accuracy of 99.4%, the improved DNN was the most accurate, followed by DNN with 98.2% and 97.5% for CNN and RNN, respectively.

Table 1 COMPARISON OF CLASSIFICATION ACCURACY RESULTS WITH THE IMPROVED DNN, DNN, RNN, AND CNN.

| epoch number | Improved DNN | DNN | RNN | CNN |
|---|---|---|---|---|
| Epoch 1 | 99.1 | 98.1 | 97.5 | 97.4 |
| Epoch 2 | 99.3 | 98.3 | 97.4 | 97.3 |
| Epoch 3 | 99.4 | 98.0 | 97.2 | 97.5 |
| Epoch 4 | 99.0 | 98.2 | 97.3 | 97.4 |
| Epoch 5 | 98.9 | 98.3 | 97.6 | 97.1 |
| Epoch 6 | 99.0 | 97.98 | 97.4 | 97.5 |
| Epoch 7 | 98.8 | 97.99 | 97.3 | 97.4 |
| Epoch 8 | 99.4 | 98.0 | 97.5 | 97.3 |
| Epoch 9 | 99.3 | 98.1 | 97.1 | 97.2 |
| Epoch 10 | 99.2 | 98.2 | 97.4 | 97.4 |

Overall, the improved DNN algorithm can be observed that attained maximum accuracy in this study's dataset. There were 100 epochs in total. Fig. 7 shows the curves of the high-accuracy model discovered by Improved DNN on the convex dataset for 100 epochs when compared to DNN, RNN, and CNN models. We can see that accuracy of our models have improvement when compared to other models, implying that the Improved DNN is actually capable of identifying perfect models for a given dataset.

## VI. Conclusion

In this paper, we have presented a novel deep learning-based model which improves DNN by applying the dropout model to classify Wheat gene expressions. In addition to, the deep learning algorithms CNN, DNN, and RNN, and the proposed model are implemented for the classification of gene expression data. Moreover, the outliers and noisy data are addressed, by using a pre-processing methodology for all features of gene expression, after that we trained all of our models individually using a perfect framework and learning method. Finally, our learned models are applied to testing data to classify it. For all of the datasets studied, the Improving-DNN outperformed other models in accuracy terms from the result illustrated our Improving-DNN has a high accuracy of 99.4%, while DNN has 98.2% accuracy, RNN and CNN have 97.5% accuracy. Therefore, the Improving-DNN model is actually more appropriate for solving the wheat gene expression dataset.

## References

[1]U.N. Desa,"World population prospects 2019: Highlights," New York (US): United Nations Department for Economic and Social Affairs, vol 11, no. 1, pp. 125. 2019.

[2]S. McGuire, "FAO, IFAD, and WFP. The state of food insecurity in the world 2015: meeting the 2015 international hunger targets: taking stock of uneven progress. Rome: FAO," Advances in Nutrition, vol. 6, no. 5, pp. 623-624, 2015.

[3]M.W.Rosegrant, S.Tokgoz, P.Bhandary and S. Msangi, "Looking Ahead: Scenarios for the Future of Food. 2012 Global Food Policy Report. IFPRI. Washington," International Food Policy Research Institute (IFPRI), vol. 4, no. 3, pp. 1-15, 2013.

[4]D. Tilman, C. Balzer, J. Hill and B. L. Befort, "Global food demand and the sustainable intensification of agriculture," Proceedings of the National Academy of Sciences - PNAS, vol. 108, no. 50, pp. 20260–20264, 2011.

[5]D. K. Ray, N. D. Mueller, P. C. West and J. A. Foley, “Yield trends are insufficient to double global crop production by 2050,” PlOS ONE, vol. 8, no. 6, pp. e66428–e66428, 2013.

[6]H. Spiertz, “Avenues to meet food security. The role of agronomy on solving complexity in food production and resource use,” European journal of agronomy, vol. 43, no. 5, pp. 1–8, 2012.

[7]J.L. Araus, R. Park, D. Calderini, D. Miralles, T. Shen et al., “Prospects of doubling global wheat yields,” Food and energy security, vol. 2, no. 1, pp. 34–48, 2013.

[8]W.Hamdy, A. Ismail, W. A. Awad, A.H. Ibrahim, and A. Hassanien, "A Support Vector Machine Model for Rice (Oryza sativa L.) Leaf Diseases Based on Particle Swarm Optimization." In Artificial Intelligence: A Real Opportunity in the Food Industry, Springer, Cham, pp. 45-54, 2023.

[9]Y. LeCun, Y. Bengio and G. Hinton, “Deep learning,” Nature (London), vol. 521, no. 7553, pp. 436–444, 2015.

[10] A. Elaraby, W. Hamdy and M. Alruwaili, “Optimization of deep learning model for plant disease detection using particle swarm optimizer,” Computers, materials & continua, vol. 71, no. 2, pp. 4019-4031, 2022.

[11] J. R. Ubbens and I. Stavness, “Corrigendum: deep plant phenomics: A deep learning platform for complex plant phenotyping tasks,” Frontiers in plant science, vol. 8, no. 12, pp. 2245–2245, 2018.

[12] A. Elaraby, W. Hamdy and S. Alanazi, “Classification of citrus diseases using optimization deep learning approach,” Computational intelligence and neuroscience, vol. 2022, no. 10, pp. 9153207–9153212, 2022.

[13] N. Ni and S. Xu, “Model optimization strategies based on deep neural networks Learning and application of pruning optimization algorithms,” Journal of Physics, Conference series, vol. 2303, no. 1, pp. 012033, 2022.

[14] A. N. Diaye, B. Byrns, A.T. Cory, K.T. Nilsen, S. Walkowiak et al., “Machine learning analyses of methylation profiles uncovers tissue-specific gene expression patterns in wheat,” The plant genome, vol. 13, no. 2, pp. e20027, 2020.

[15] K. S. Sandhu, D. N. Lozada, Z. Zhang, M. O. Pumphrey and A. H. Carter, “Deep learning for predicting complex traits in spring wheat breeding program,” Frontiers in plant science, vol. 11, no. 4, pp. 613325–613335, 2021.

[16] L. Li, M.A. Hassan, S. Yang, F. Jing, M. Yang et al., "Development of image-based wheat spike counter through a Faster R-CNN algorithm and application for genetic studies." The Crop Journal, vol. 12, no. 3, pp. 1-12, 2022.

[17] P. Matteo, "Machines that morph logic: neural networks and the distorted automation of intelligence as statistical inference." Glass Bead, vol. 1, no.1, pp. 25-36, 2017.

[18] L. B. Balzer and M. L. Petersen, “Invited commentary: machine learning in causal inference-how do I love thee? let me count the ways,” American journal of epidemiology, vol. 190, no. 8, pp. 1483–1487, 2021.

[19] I. H. Sarker, “Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions,” SN computer science, vol. 2, no. 6, pp. 420–420, 2021.

[20] G. Jing, P. Li, Z. Chen and J. Zhang, "A survey on deep learning for multimodal data fusion," Neural Computation, vol. 32, no. 5, pp. 829-864, 2020.

[21] G. R. T. de Lima and G. B. Scofield, “Feasibility study on operational use of neural networks in a flash flood early warning system,” Revista Brasileira de Recursos hídricos, vol. 26, no. 2, pp. 1-10, 2021.

[22] J. M. Silva, A. Figueiredo, J. Cunha, J.E. Dias, S. Silva et al., “Using rapid chlorophyll fluorescence transients to classify vitis genotypes,” Plants (Basel), vol. 9, no. 2, pp. 174-189, 2020.

[23] M. Mostavi, Y.-C. Chiu, Y. Huang and Y. Chen, “Convolutional neural network models for cancer type prediction based on gene expression,” BMC medical genomics, vol. 13, no.5, pp. 1–13, 2020.

[24] S.D. O'Donovan, K. essens, D. Lopatta, F. Wimmenauer, A. Lukas et al., “Use of deep learning methods to translate drug-induced gene expression changes from rat to human primary hepatocytes,” Plos One, vol. 15, no. 8, pp. e0236392, 2020.

[25] H. Lahmer, A. E. Oueslati and Z. Lachiri, “Classification of DNA microarrays using deep learning to identify cell cycle regulated genes,” 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), pp. 1–5, 2020.

[26] B. He, L. Bergenstrahle, L. Stenbeck, A. Abid, A. Andersson, et al., “Integrating spatial gene expression and breast tumour morphology via deep learning,” Nature biomedical engineering, vol. 4, no. 8, pp. 827–834, 2020.

[27] O. Ahmed and A. Brifcani, “Gene expression classification based on deep learning,” 4th Scientific International Conference Najaf (SICN), Najaf, Iraq, pp. 145–149, 2019

[28] K.W. Jordan, S. Wang, F. He, S. Chao, Y. Lun et al., “The genetic architecture of genome‐wide recombination rate variation in allopolyploid wheat revealed by nested association mapping,” The Plant journal: for cell and molecular biology, vol. 95, no. 6, pp. 1039–1054, 2018.

# Multi Agent Reinforcement Learning based Feature Selection Algorithm for Class Imbalance Problem: An Application to Prediction of Postoperative Outcomes in End-stage Renal Disease Patients

SeoHee Kim[1*], SunYoung Park[2], and Jiyoung Woo[1]
[1]*ICT Convergence Soonchunhyang University Asan, Republic of Korea*
[2] *Anesthesiology and Pain Medicine Soonchunhyang University Seoul Hospital Seoul, Republic of Korea*
*Contact: jywoo@sch.ac.kr

*Abstract*— **In predicting postoperative outcomes for patients with end-stage renal disease, our study encounters challenges related to class imbalance and a high-dimensional feature space. Therefore, we propose a novel approach to feature selection using multi agent reinforcement learning (MARL).**
**In contrast to previous research, we firstly set up the deterministic reward using the Mutual Information (MI) of variables with a target class to select the variables that have high dependence with a class variable. Then, we prioritize variables that influence the minority class using Shapley Additive Explanations (Shap) values to enhance interpretability. Finally, we construct a comprehensive reward framework by combining the dynamic reward of the classification model performance and Shap values and the static reward of Mutual Information (MI) to enhance performance and reduce dimensionality simultaneously. Our approach successfully solved the class imbalance problem and achieved the best performance from the ablation analysis. Our model achieved a 16\% increase in the F1 score for the minority class and an 8.22\% increase in the overall F1 score compared to the baseline Extreme Gradient Boosting (XGBoost) model.**

## I. Introduction

Multi-agent reinforcement learning (MARL) [1] is a framework where multiple agents learn how to interact and cooperate in a shared environment to achieve common goals. In this study, we propose a feature selection model with a novel reward model, which is working in MARL for identifying optimal feature combinations. We applied our proposed model onto data collected from patients with end-stage renal disease to predict the postoperative outcomes. Therefore, we will develop a model to predict Major Adverse Cardiovascular Events (MACE) after surgery using patients' surgical data, anaesthesia records, and Electronic Medical Records (EMR) data. Given the high dimensionality of the generated data based on this dataset, feature selection was necessary to reduce model complexity and computational costs. Therefore, we designed MARL with a novel reward function in this study to discover the optimal feature combination that enhances the model performance and reduces the model complexity.

Our proposed reward function enhances interpretability while reducing the dimensionality of the model using Mutual Information (MI) [2] and the Shapley Additive Explanations (SHAP) [3] algorithm. Additionally, in this study, we solved the class imbalance issue by developing a model selecting variables that hold significant influence over the minority class, aiming to predict the minority class better. As a result, this study contributes important research findings that multi-agent based feature selection method can be a promising approach for solving the class imbalance problem.

## II. Related Work

We surveyed the literature on RL based variable combinations to gain insights into discovering optimal variable combinations in multidimensional data. Through this process, we aim to explain our research objectives and direction.

Fan et al. [4] proposed an interactive RL framework for feature selection. This framework advised a hesitating agent by utilizing the importance of variable using K-best and Random Forest (RF) algorithm. Over time, the agent gains self-learning capabilities through exploration and learning without a trainer. Kim et al. [5] proposed a MARL-based feature selection algorithm. This algorithm distinguishes the roles of guide agents and main agents, providing rewards for the main agent's actions and enabling it to evaluate the validity of its actions. The guide agent makes random decisions independent of prior experience. By comparing the actions of these two agents, experience is integrated into the learning process, ultimately leading to improved feature selection.

While previous studies have overlooked several crucial aspects. Firstly, previous research has not addressed the reward distribution approach considering variables that influence the minor class. To address this issue, we introduce a novel reward model that assigns more rewards to variables influencing the minor class alongside the feature selection process.

Furthermore, in prior research, there was a limited research on explainable feature selection in MARL-based feature selection algorithm. Therefore, in this study, we took a novel approach by incorporating shap into the reward of the RL model. Through this approach, we successfully developed an explainable MARL model that utilizes shap values to extract the influence of variables.

## III. Model

### *A. Feature Selection Using MARL*

Feature selection alleviates the issue caused by the curse of dimensionality, prevents overfitting, and enhances the generalization performance of models. Given the use of 250 variables in this study, it is essential to narrow down the selection to those that substantially impact performance. Additionally, due to the significant data imbalance, it is crucial to consider variables that positively impact the minority class when seeking variable combinations. While there are various feature selection methods, this study employed MARL to consider optimal combinations. Defining the critical components of MARL namely the agent, action, and reward, holds paramount importance as they directly influence the learning algorithm's performance.

1) Agent

In RL, an agent is a decision-maker that selects actions within a given state. Specifically, the agent serves as the entity for performing variable selection, creating 250 agents to match the number of features. These agents are denoted as $AG_i$, where $i$ ranges from 1 to 250.

2) Action

In RL, an action refers to an agent's choice while interacting. This study defines actions as deciding whether to include a specific variable for feature selection. An action is represented by either 0 or 1. A value of 0 signifies the non-selection of a corresponding variable, while a value of 1 signifies the selection of that variable. The current action is $A_t$, whereas the combination of past actions is $A_{t-1}$. This encompasses all actions selected by $AG_i$. A comparison between the current and past states is conducted to assign rewards. In the current state, actions chosen by each $AG_i$ are denoted as $A_{i,t}$, while actions selected in the previous state are represented as $A_{i,t-1}$.

3) Reward

A reward represents the feedback an agent receives when interacting with its environment, reflecting the evaluation or performance of the agent's actions within a specific state. In this experiment, the overall F1-score of the Extreme Gradient Boosting (XGBoost) [6] classification model is defined as the major reward.

The Rewards were assigned to the $A_t$ by comparing the F1-score obtained from two actions in previous and current episodes. If the current F1 performance, denoted as $F_t$, is greater than the previous reward, $F_{t-1}$, the reward system is adjusted positively; otherwise, it is adjusted negatively. Rewards are assigned only when two consecutive actions, $A_t$ and $A_{t-1}$, are different and the reward is distributed by dividing it according to the number of agent that changed the action, denoted as n. This difference of the classification model performance according to the current and past actions is denoted as $F_{delta}$, and its formulation is provided in (1).

$$F_{delta} = \frac{F_t - F_{t-1}}{|n|}, n = \in \{n | A_{i,t} \neq A_{i,t-1}\} \quad (1)$$

Therefore, Assigning the same reward to actions with differences between the $A_t$ and $A_{t-1}$ is inappropriate. To address this issue, we proposed to assign more rewards to variables closely related to the target-dependent variable using the MI, denoted as $W_{MI}$.

Moreover, We will utilize shap values, denoted as $W_{Shap}$, to assign greater rewards to variables that positively explanatory power onto the minority class. As shown in (2), we add two rewards, $W_{MI}$ and $W_{Shap}$ that reflect the individual rewards of variable based on contribution of feature reduction and class imbalance problem. The proposed reward function appropriately considers variables sensitive to performance enhancement and class imbalance while optimizing variable combinations.

$$R_t = F_{delta} + W_{MI} + W_{Shap} \quad (2)$$

MI [2] is a metric that quantifies the dependency and relationship between two variables, representing the amount of information that one probability variable provides about another. A higher MI indicates that the two variables are associated with each other, and the values of one variable offer useful information about of the other variable. Shap [3] algorithm estimates the contributions of each input variable by considering interactions among variables. For that, the algorithm computes the average of shapley values across various variable combinations. We employ the shap value to put emphasis on the predictive outcome of each variable in the minority class for reward function definition.

## IV. Experiment

Data were obtained from 1,632 patients from March 2018 to April 2020. The dataset comprises EMR and anesthesia records, including a total of 250 variables. The pre-operative EMR features include demographic information, lab results, check-ups, and medical conditions. Anesthesia records are transformed into dummy variables. Vital signs like blood pressure, heart rate, and oxygen saturation are organized into descriptive and peak features. In this dataset, "MACE" accounts for only 5.5%, which is significantly lower than "Non-MACE", resulting in a class imbalance issue. In such cases, the model tends to be biased towards the majority class, making it challenging to learn from the minority class accurately.

Therefore, In this study, we utilized random sampling, a method of forming a sample by randomly selecting individuals from the population. Additionally, we conducted classification using the xgboost algorithm. This model is designed to perform more accurate predictions, preventing overfitting and providing rapid learning. We employed this approach in our study.

(7) represents the process of updating the Q-value function in MARL. In this equation, $Q(a)$ represents the current Q-value for a specific action a, while $R_t$ denotes the reward received at time $t$. The parameter $\alpha$ signifies the learning rate, determining how much the current value is adjusted based on the difference between the previous value and the current reward. For our proposed feature selection model, we set $\varepsilon$ at a reduced value of 0.05 given the considerable number of variables. This decision aimed to facilitate a broader exploration of actions while accumulating experience. Furthermore, $\alpha$ was set to 0.2. The decay rate for reducing the probability of randomly exploring actions is denoted as $\varepsilon$. The decay rate for facilitating the learning rate, denoted as $\alpha$ decay rate, was both set to 0.995. Additionally, the initial value of Q-value was set to -1 for conducting the experiments.

$$Q(a) = Q(a) + \alpha * \big(R_t - Q(a)\big) \quad (7)$$

## V. Result

Considering the data's imbalanced nature, the evaluation of performance employed F1-score as alternatives to accuracy. The XGBoost model, which achieved the highest performance, was used to build the predictive model, reaching an F1 score of 81.3% even without feature selection. Additionally, the boruta [7] algorithm, a RF-based variable selection method, was employed. As a result, 62 variables were selected, and the performance improved to 82.6%. However, instead of considering variable importance-based selection, we explored the application of MARL to find optimal variable combinations.

Building upon the experiment by Kim et al [5], the outcome indicated a performance enhancement to 83.7% compared to the case without RL. However, the F1-score for the minor class remained lower at 69.0%. Focusing solely on the comparison between past and present agent actions, an F1-score of 82.53% was achieved. Notably, Significant MI values with weights led to an 85.08% performance improvement. Applying shap values to all classes yielded an F1 score of 86.48%. For the minor class, shap values resulted in a performance of 87.37%. This result emphasizes the potential to create a more accurate model by assigning shap value weights to the variables influencing the minor class. Furthermore, we conducted experiments incorporating MI and shap values to assign weights. In this case, when conducting experiments by exclusively assigning rewards to variables that influence the minor class rather than constructing rewards based on shap values that impact the entire class set, we attained the highest performance with an F1-score of 89.52%. Notably, this achievement extends to the minor class with an F1-score of 80.00%. This observation implies the effective extraction of variables that influence the minor class. The experimental findings are consistent with Table 1.

TABLE I
EXPERIMENT RESULT TABLE

| Model | $W_{Shap}$ | F1 for Non-MACE | F1 for MACE | F1 | Number of Features |
|---|---|---|---|---|---|
| XGBoost | X | 98.6 | 64.0 | 81.3 | 250 |
| Boruta | X | 98.6 | 66.7 | 82.6 | 62 |
| MARFS[5] | X | 98.6 | 69.0 | 83.7 | 87 |
| RL | X | 98.4 | 66.67 | 82.53 | 72 |
| MI RL | X | 98.72 | 71.43 | 85.08 | 78 |
| Shap RL | All | 98.88 | 74.07 | 86.48 | 90 |
| Shap RL | Minor | 98.88 | 75.86 | 87.37 | 78 |
| MI + Shap RL | All | 98.72 | 71.43 | 85.08 | 88 |
| MI + Shap RL | Minor | 99.04 | 80.00 | 89.52 | 87 |

*All : Variables influencing the entire class / Minor : Variables influencing the minor class

The shap results confirm that the administration of dopamine, epinephrine and the operation department, are a significant predictive factors for MACE. Summarizing the MI analysis results, it was observed that the descriptive features and peak features from vital signs, used for intraoperative monitoring, exhibited higher MI values. Furthermore, We analyzed the variable importance of xgboost to identify the variables that played a significant role in improving the performance. we also observed that variables such as phenylephrine and Colloid impacted the improvement of the Minor class performance. However, variables such as surgical risk score and physical status were selected when influencing the entire set of variables, but were excluded otherwise. This indicates that these two variables do not significantly affect the Minor class, and their removal can even lead to enhanced performance.

## VI. CONCLUSIONS

This study aimed to address the class imbalance problem and enhance the interpretability of MARL models for feature selection. In contrast to previous research, we introduced a reward distribution approach that assigns higher rewards to variables influencing the minor class. To achieve this, we utilized shap values to enhance the selection of variables impacting the minor class while also improving interpretability. Furthermore, through the combination of shap and MI, we crafted a comprehensive reward framework that achieved simultaneous performance enhancement and dimension reduction. This strategic approach allowed us to enhance the performance of the minor class, ultimately optimizing overall performance. However, this study did not consider the importance of variables in the medical context. Therefore, future research will focus on excluding or including medically irrelevant variables to enhance the model and plan for additional investigations.

## VII. ACKNOWLEDGEMENT

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ICAN (ICT Challenge and Advanced Network of HRD) program (IITP-2023-2020-0-01832) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation)

## REFERENCES

[1] Richard S Sutton and Andrew G Barto. Reinforcement learning: An introduction. MIT press, 2018.

[2] Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and Devon Hjelm. Mutual information neural estimation. In International conference on machine learning, pages 531–540. PMLR, 2018.

[3] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, Advances in Neural Information Processing Systems, vol 30. Curran Associates, Inc., 2017.

[4] Wei Fan, Kunpeng Liu, Hao Liu, Pengyang Wang, Yong Ge, and Yanjie Fu. AutoFS: Automated feature selection via diversity-aware interactive reinforcement learning. In IEEE International Conference on Data Mining (ICDM), pages 1008–1013, 2020.

[5] Minwoo Kim, Jinhee Bae, Bohyun Wang, Hansol Ko, and Joon S. Lim. Feature selection method using multi-agent reinforcement learning based on guide agents. 23(1):98, 2023.

[6] Tianqi Chen, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, Kailong Chen, Rory Mitchell, Ignacio Cano, Tianyi Zhou et al. Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4):1–4, 2015.

[7] Miron B Kursa and Witold R Rudnicki. Feature selection with the boruta package. Journal of statistical software, 36:1–13, 20

# Moving Object Tracking on a Smartphone using a Cradle Head Servo Motor

Neunggyu Han [1], Sun Joo Ryu [2], Seungmin Rho[3], and Yunyoung Nam [4]
[1] *Department of ICT Convergence, Soonchunhyang, Asan, 31538, Republic of Korea*
[2] *Department of Enterprise School, Soonchunhyang University, Asan 31538, Korea of Korea*
[3]*Department of Industrial Security, Chung-Ang University, Seoul 06974, South Korea*
[4] *Department of Computer Science and Engineering, Soonchunhyang, Asan, 31538, Republic of Korea*

*Abstract*— **The increasing demand for artificially intelligent smartphone cradles has prompted the need for real-time moving object detection. Real-time moving object tracking requires developing algorithms for instant analysis of tracking without delays. Especially developing the system on smart phones should be considered the different operating systems and software development environments. Issues of current technology of real time moving object tracking systems arise when small and large objects coexist, causing the algorithm to prioritize larger objects or struggle with consistent tracking across varying scales. Fast object motion further complicates accurate tracking, leading to potential errors and misidentification. To address these issues, we propose deep-learning-based real time moving object tracking systems. For object detection, two approaches are proposed. The first approach, called as the accuracy priority mode, achieves a balance between high accuracy and speed required in the smartphone environment. The second approach, the speed priority mode, highlights optimizing the speed of the inference rates in order to track fast moving objects. The accuracy priority mode incorporates CSPNet with ResNet to maintain the high accuracy, whereas the speed priority mode simplifies the convolutional layer complexity while maintaining accuracy. In experiments, we evaluated both modes in terms of accuracy and speed.**

**Keywords:** real-time moving object tracking; deep-learning; cradle with smartphone; cradle servo motor

## I. INTRODUCTION

Due to the continuous improvement in smartphone performance and the widespread use of deep learning technology, object detection techniques based on smartphones using a cradle servo motor have gained attention from researchers [1-4]. Real-time moving object detection technology has been widely used in various fields such as robotics, transportation, manufacturing, and security. Especially, with the continuous growth of the creator economy, the total addressable market is projected to reach $480 billion by 2027 [5]. As the number of YouTube creators and the size of its market increase, the demand for advanced technology to support content creation also grows. YouTube creators mainly rely on an artificial intelligent smartphone cradles for content creation. The real-time moving object detection technology has undergone rapid development due to its inherent portability and user-friendly nature.

Hwang and Liao developed a system using a moving camera with a servo cradle to enable a robot to track and imitate human motions [1]. The system achieves real-time pose imitation of 3D human motions by separating and tracking the upper body (UB) and lower body (LB) movements, and then recombining them during the tracking process. The servo cradle-head RGB-D vision system (SCH-RGB-D-VS) is used for capturing the 3D motions of the target human (TH) for imitation [1]. Servo motors control the movement and orientation of the camera or sensors, ensuring that the target human's motions are correctly captured within the field of view (FOV) of the vision system. This allows the humanoid robot to imitate the poses and movements of the target human in real-time.

Asali et el. presented a real-time video tracking system for visual servo applications using Support Vector Machines (SVM) [2]. The system aims to maintain a target object's position in the camera field of view by tracking it in complex environments and lighting conditions. It combines a Sony FCB-EV7520 camera and the STRUCK algorithm for tracking, while a Proportional-Integral-Derivative (PID) controller is used for visual servo control. Experimental results demonstrate the system's effectiveness in real-time tracking of moving objects in indoor and outdoor settings.

Zhang Shu et al. presents an automatic camera servo system using an improved frame differential algorithm to detect and track moving objects in real-time [3]. The system analyzes video streams, sets appropriate thresholds to distinguish noise from movement, and sends control orders to the cradle head via the serial port for tracking. It performs well in various circumstances and incorporates video capture, temporal matching, video preservation, compression, and auto-alarm functions for monitoring.

However, current technology of real time moving object detection systems on smartphones faces several limitations. If there are small and large objects present simultaneously, the algorithm may prioritize larger objects or struggle to maintain consistent tracking across objects of varying scales. It struggles with accurately detecting and tracking objects when they are occluded or overlapping with each other. In scenarios where multiple objects are close together or obstructed, the algorithm stops tracking or has a tendency to track only the new object appeared later in the camera on the smart phone. The cradle's performance is also affected when objects are in fast motion in smartphone. Quick movements can make it challenging for the algorithm to accurately track the objects, leading to potential tracking errors or misidentification.

In addition, the development of moving object tracking technology using cradle servos has been mainly focused on

computer-based research. Studying and implementing this technology in a smartphone environment presents some differences and challenges with hardware limitations and computational resources availability. Smartphone operating systems and software development also have several restrictions and requirements, which can add complexity to the integration of cradle servo technology.

Utilizing deep learning principles, this study has led to the creation of algorithms designed to address object recognition, real-time object detection, and tracking. This involves the application of Shape, Color, and Motion-based Classification techniques to categorize objects based on their distinctive attributes. Overcoming the challenge of tracking multiple objects, particularly when overlaps occur, is achieved through the incorporation of trajectory and velocity data from the classified entities. This approach guarantees seamless and continuous tracking, with a steadfast focus on the intended target. Additionally, our algorithms exhibit the capability to discern noteworthy actions within uninterrupted motion sequences, consequently resulting in an improved level of tracking precision.

Two approaches are proposed in this paper for object detection. First approach is called as accuracy priority mode with a focus on high accuracy detection and tracking. This mode incorporates CSPNet [4], which employs residual blocks, with ResNet [5], a structure that divides the input into two branches. One branch passes through the network, while the other branch is concatenated with the resultant features. The second approach is the speed priority mode, which prioritizes a fast inference rate. This mode is specifically designed for rapid processing and achieves a notable speed improvement. The speed priority mode reduced the number of layers, while maintaining a commendable level of accuracy. Within a single application, users can choose between these two approaches based on their priorities. Opting for the first approach enables users to prioritize accuracy, making it suitable for tasks such as accurate differentiation between dogs and cats. However, in scenarios involving rapidly moving objects, the first approach might prove inadequate, making the second approach a more suitable choice.

The object tracking process involves the cradle servo continuously tracking objects and accurately positioning them at the camera's center. This is achieved by calculating the object's velocity through the assessment of positional changes between consecutive frames, as well as measuring the spatial distance between the object and the camera's focal center.

The implementation of the two modes and cradle servo systems involved the following research process. Initially, a dataset consisting 100 YouTube videos was collected, and from these videos, specific data related to humans, dogs, and cats were extracted, resulting in a dataset of 65,000 frames. Following data extraction, pre-trained weights were applied for preliminary labeling and review. Subsequently, the accuracy priority mode was developed with a primary focus on achieving high accuracy while minimizing computation costs. The mean average precision at 50 Intersection over Union (mAP50 IoU) was adopted as the accuracy evaluation metric for performance assessment. Moreover, the mode's computational cost was quantified using Giga Floating Point Operations (GFLOPs). The mode's architecture was established by integrating elements from the ResNet structure introduced in YOLOv3 [6] and the CSPResNet utilized in YOLOv4 [7]. A comprehensive evaluation of both computation cost and accuracy was conducted to determine the optimal model. Lastly, the derived model served as the basis for creating a speed priority variant by simplifying the number of layers by testing multiple times to find the optimal layers to fit into our model.

## II. Materials and Methods

### *1. Proposed Methodology*

Figure 1 depicts the diagram of the system proposed. This system is designed to track moving objects in real-time based on the video captured using a smartphone cradle servo. The first process of object tracking involves the functionality to detect and recognize objects. To mitigate the existing issue of objects, particularly small ones, not being recognized from a distance by the camera, an accuracy priority mode was developed to ensure the recognition of even small objects. We aimed to enhance the accuracy of classification to effectively distinguish between objects such as humans, dogs, and cats, while simultaneously improving the accuracy of recognition regardless of the size of the object and its distance from the camera.

Furthermore, to enhance the tracking accuracy of rapidly moving objects, we developed a speed priority mode with a focus on optimizing processing speed. The conventional YOLO model [8] is designed to excel on the COCO dataset [9], which comprises 80 classes, presenting limitations in terms of model light weighting. The COCO dataset demands precise classification and recognition of a wide array of everyday objects across its 80 classes. Even the latest iteration, YOLOv8 [10], falls short of achieving an accuracy of 55% in terms of mAP50-95 evaluation. However, this research is primarily aimed at refining servo usage for content creators. Given that the prominent objects mainly include humans, dogs, and cats, this paper predominantly targets the recognition of these entities. Consequently, as only three classes need to be classified and recognized, there is potential for a more lightweight and streamlined approach.

The aforementioned accuracy priority mode and speed priority mode are developed to be selectable by users, allowing for mode changes as needed. Both accuracy priority mode and speed priority mode are grounded in enhancing recognition speed in the smartphone environment. Unlike a computer environment, the smartphone environment has relatively limited computing performance, necessitating the lightweighting of detection models to ensure real-time tracking. Typically, a processing speed of at least 10 frames per second is required, driving the development of both modes with a focus on lightweighting to achieve this speed requirement.

Next, the extracted feature maps of detected objects are compared for similarity to assign IDs. During the process of extracting object feature maps, data from the Region of Interest (ROI) of the recognized object is obtained. This data passes through a CNN-based feature extractor to be transformed into feature maps. This CNN-based feature extractor is composed of CNN layers trained on humans, dogs, and cats. In the ID

assignment process, based on the feature maps, the features of the object in the current frame are compared with those of the previous frame. If the similarity between the existing feature map and the current one exceeds 50%, the same ID is reassigned. Otherwise, a new ID is assigned.

Subsequently, the design of the smartphone cradle servo was tailored to follow the identification tag of the foremost recognized object from within a pool of multiple detected IDs. This engineering approach aims to mitigate tracking inaccuracies stemming from the simultaneous presence of diverse, both sizable and petite, objects. The primary users of the cradle servo are predominantly content creators, and in a majority of instances, the earliest-appearing object assumes the role of the main subject. Consequently, the servo's design prioritizes a tracking paradigm rooted in the order of recognition, independent of the object's dimensions, with this particular use case taking precedence. While the possibility to selectively track a designated object in isolation is viable, for experimental purposes, a simplified iteration was incorporated.

Moreover, a deliberate effort was made to align the tracked object's position with the central axis of the camera. This alignment was established through a process involving the computation of the measurement of its spatial separation from the camera's focal center. The methodology employed encompassed the strategic adjustment of the object's coordinates, facilitating its relocation to the exact center of the camera's display. This procedural approach substantiates the consistent maintenance of the object's central alignment with respect to the camera's viewpoint.

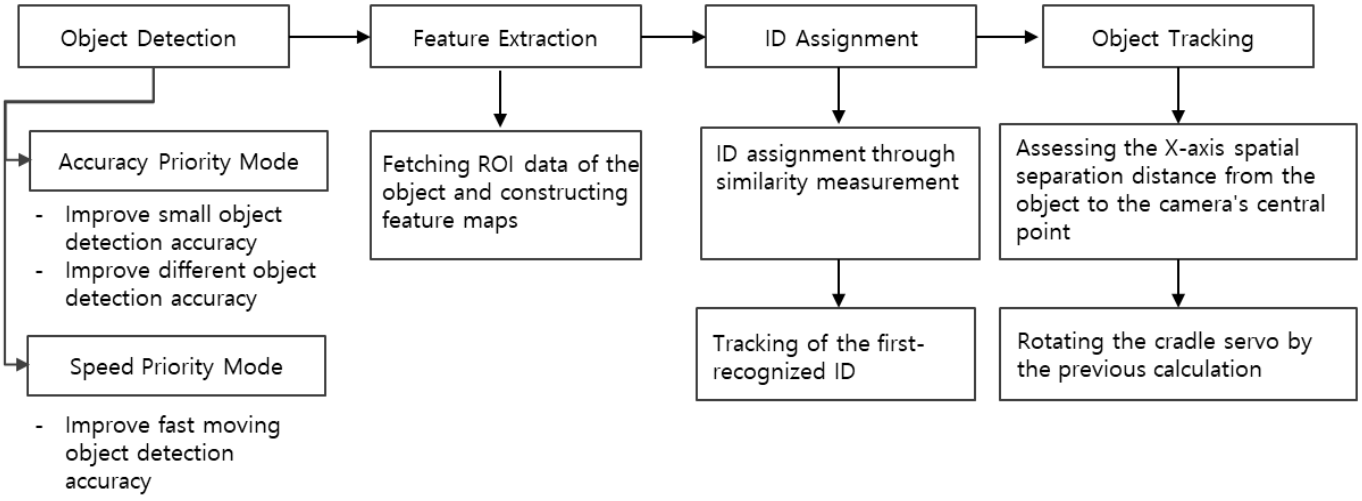



**Figure 1.** Main flow diagram of the proposed methodology

*2. Environmental Setup*

The experimental setup encompassed two distinct environments. The training environment is denoted as Table 1 and the implementation environment on smartphone is represented in Table 2.

The training of the YOLO model was carried out within an environment described in Table 1. The setup of the environment involved Docker, wherein the nvidia/cuda:11.6.2-cudnn8-devel-ubuntu20.04 image was selected from a range of Docker images. This specific image encompasses both NVIDIA's CUDA Toolkit and cuDNN. The environment was used as the backdrop for augmenting the capabilities of the existing YOLOv8 platform. This augmentation was achieved through adjustments tailored to the construction of desired models. It's noteworthy that the original platform lacked specific essential layers, like "Shortcut," pivotal for configuring novel model architectures. Hence, recourse was taken to a modified platform, which featured the requisite additional layers, thus facilitating the training of models.

**Table 1.** Train Environment

| Train Environment | |
|---|---|
| CPU | Intel Xeon Silver 4216 x2 |
| RAM | 192GB |
| GPU | RTX A5000 x2 |
| OS | Ubuntu 20.04 |
| YOLO Platform | Modified YOLOv8 |

For the operating system in experimental stage, Android 13 was employed. YOLOv8 was used for object recognition and tracking algorithms. YOLOv8 employs an innovative repository to support both object detection and instance segmentation. It provides an integrated weight converter, contributing to enhanced portability across various libraries. Its architectural structure is characterized by a relatively shallow design, setting it apart from alternative versions and enabling uncomplicated customization. YOLOv8 employs an anchor-free model that directly anticipates the object's center, diverging from the use of anchor box offsets. Consequently, this approach accelerates the non-maximum suppression (NMS) process.

TensorFlow Lite was used for the inference framework in implement the smartphone application. YOLOv8 is typically implemented by using Torch. Torch-based smartphone frameworks have a performance bottleneck due to the lack of

GPU support. TensorFlow Lite, on the other hand, allows easy GPU acceleration and provides example code, making it a more efficient choice for mobile applications. We converted YOLOv8 by implementing TensorFlow Lite instead of using Torch.

**Table 2.** Implementation Environment on Smartphone

| Smartphone and Cradle | |
|---|---|
| Smartphone Model | GalaxyS21+ |
| OS | Android 13 |
| Inference Framework | TensorFlow Lite 2.8.0 |
| Cradle Model | Pivo Pod Silver |

YouTube video data was utilized as our primary dataset due to its availability and diverse conditions, including composition, angle, and brightness. To ensure an extensive and representative dataset, a total of 100 videos were collected, encompassing various resolutions ranging from HD (1280x720) to UHD (3840x2160). As the YOLOv8 platform primarily uses images for training rather than video, frame data was extracted from the collected videos at a rate of 2 frames per second (fps) to create image datasets for training. The extracted frames were then subjected to a labeling process before being utilized for training purposes.

The labeling process involved automated labeling using pre-trained weights, followed by a validation stage. Automated labeling employed the pre-trained weights of the YOLOv8x,

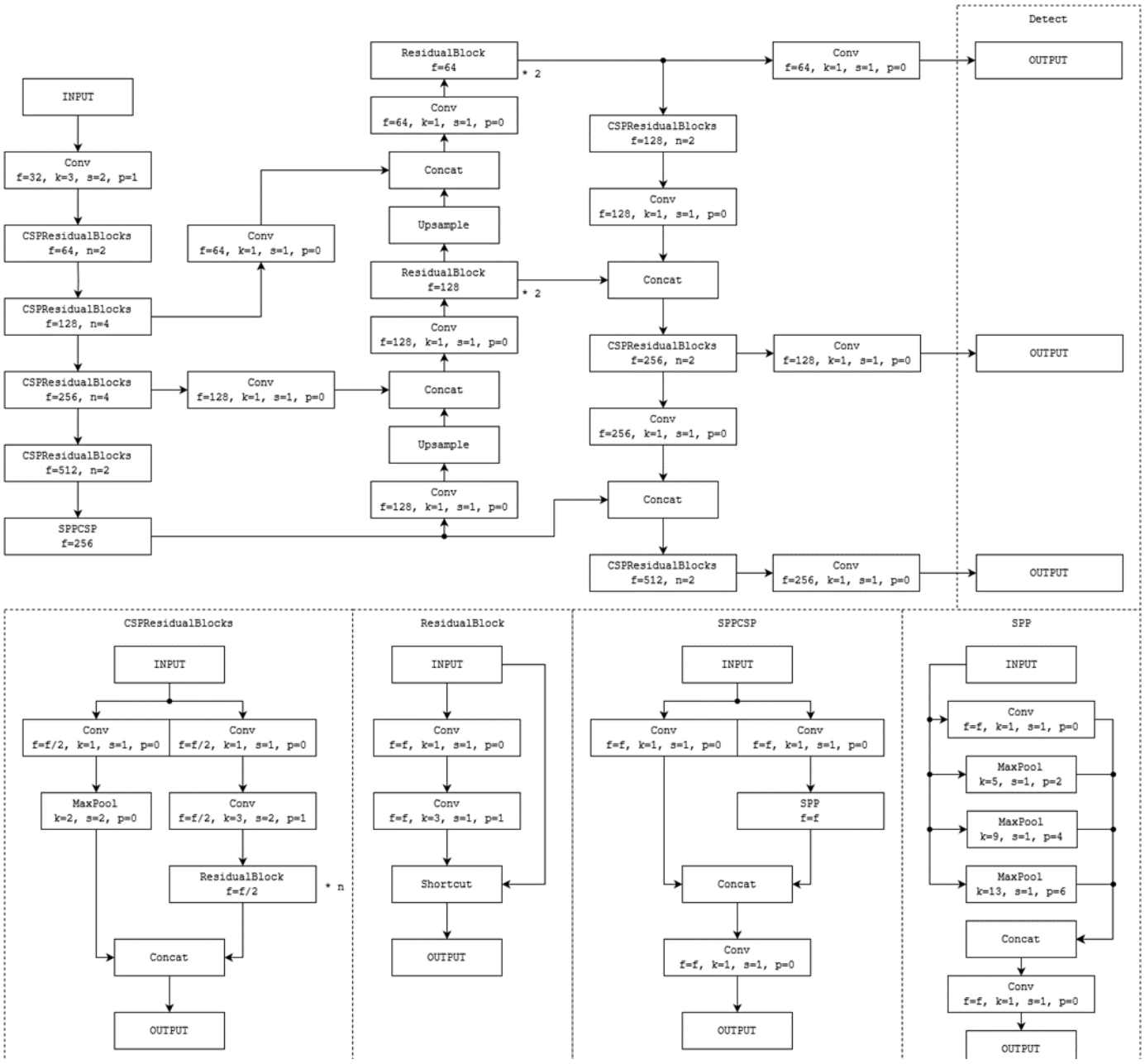



**Figure 2.** Diagram of the accuracy priority mode

divided into an inference stage and a post-processing stage. During the inference stage, the pre-trained weights of the resulting in elevated inference accuracy in comparison to standard FPN utilization.

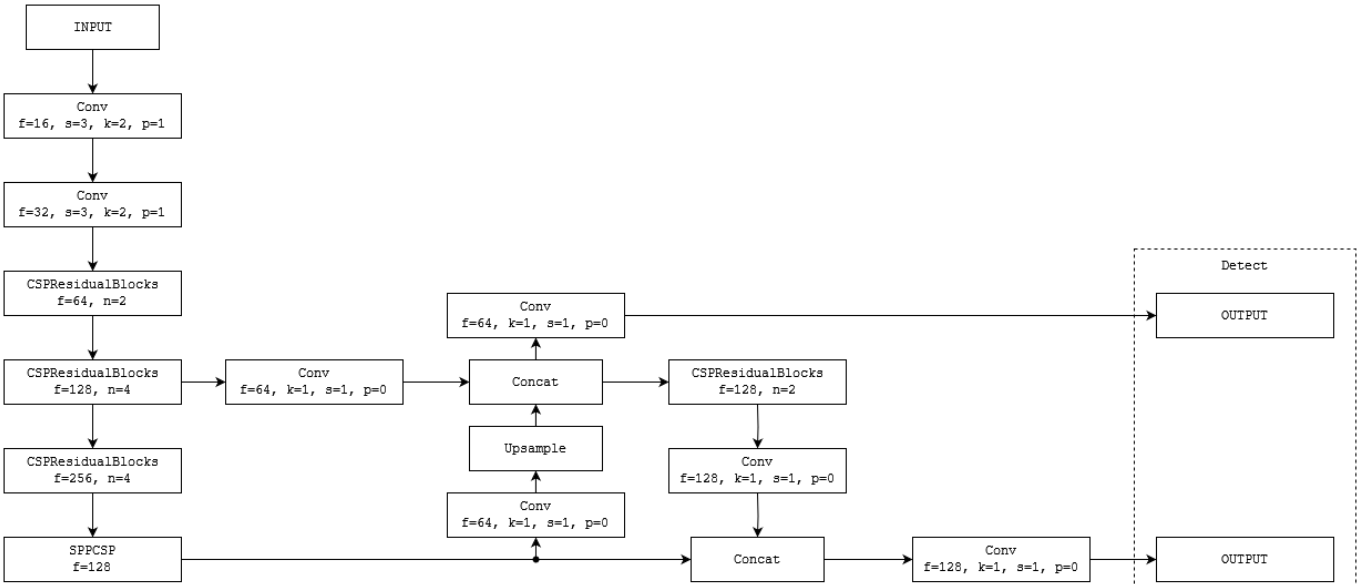



**Figure 3.** Diagram of the speed priority mode

YOLOv8x were employed to infer results for the images in the dataset. Given that YOLOv8x is trained on 80 classes, results are obtained for all 80 classes. From these results, the process of interest involves selecting and computing the coordinate values for the desired classes, which then form the basis for the labeling, constituting the post-processing stage. In this post-processing phase, all data except for humans, dogs, and cats were discarded, and subsequent steps involved the computation of the central coordinates along the x and y axes, as well as the box dimensions. This culminated in the generation of label data.

Following the automated labeling process, a subsequent manual validation step is performed for each individual label dataset. During this phase, corrective actions were taken in instances where erroneous object labeling or incorrect class assignments occurred. For instance, cases where ordinary garments were erroneously labeled as humans, or where dogs were mistakenly labeled as cats, were encountered. In such scenarios, respective label data was edited by either removal or class name adjustment. This comprehensive procedure was applied across a dataset comprising 100 videos. Upon completion of the labeling process, 90 of these video datasets were employed as the training dataset, while the remaining 10 were utilized as the validation dataset.

*3. Mode Configuration for Object Detection*

The object detection framework utilizes two distinct modes: an accuracy priority mode and a speed priority mode. The accuracy priority mode was designed to achieve swifter performance than the YOLOv8s model while minimizing potential reduction in accuracy. As visually presented in Figure 2, it derives its foundation from the Path Aggregation Network (PANet) [11], which serves as an extension of the Feature Pyramid Network (FPN) structure [12] deployed within the YOLOv8 context. This extension entails an added upscaling step from the downscaled outcomes of FPN,

Moreover, this mode integrates a block rooted in the application of the CSPNet to the ResNet. This block is constructed through an adapted rendition of the CSPNet structure. While CSPNet divides input data into two segments, linking one part to the existing network and concatenating the output with the other part, the specified block also segments inputs into two divisions. One division connects to down-sampling convolutional layers and the existing network, while the other interfaces with max-pooling layers. The outcomes from both divisions are subsequently combined. This structure combines Residual Blocks inspired by ResNet to form CSPResidualBlocks.

In the neck segment which serves as the intermediary between the Backbone and the Head, the mode commences with SPPCSP (Spatial Pyramid Pool; SPP [13] with CSPNet) and shifts to employing ResidualBlocks instead of CSPResidualBlocks during the upscaling expansion. This modification is driven by the limitation of CSPResidualBlocks within this mode to effectively accommodate the disparity between input and output sizes. As the mode advances toward the upscaling phase, there is a reduction in the number of filters. The CSPResNet architecture detailed in this study entails a twofold increase in filter count relative to the input, which presents a challenge when attempting to match or halve the output size while maintaining a consistent input. Such an adjustment would result in an overall decrease in filters across the structure, possibly leading to the loss of important features. Consequently, upon the completion of the upscaling phase, the subsequent downscaling expansion incorporates CSPResidualBlocks. During this stage, the augmented filter count facilitates the effective application of this architectural structure.

The model designed for speed priority mode is characterized by a commendably fast inference speed without significantly compromising accuracy. This particular mode streamlines its architecture by reducing the number of

extraction layers and filters compared to the accuracy mode

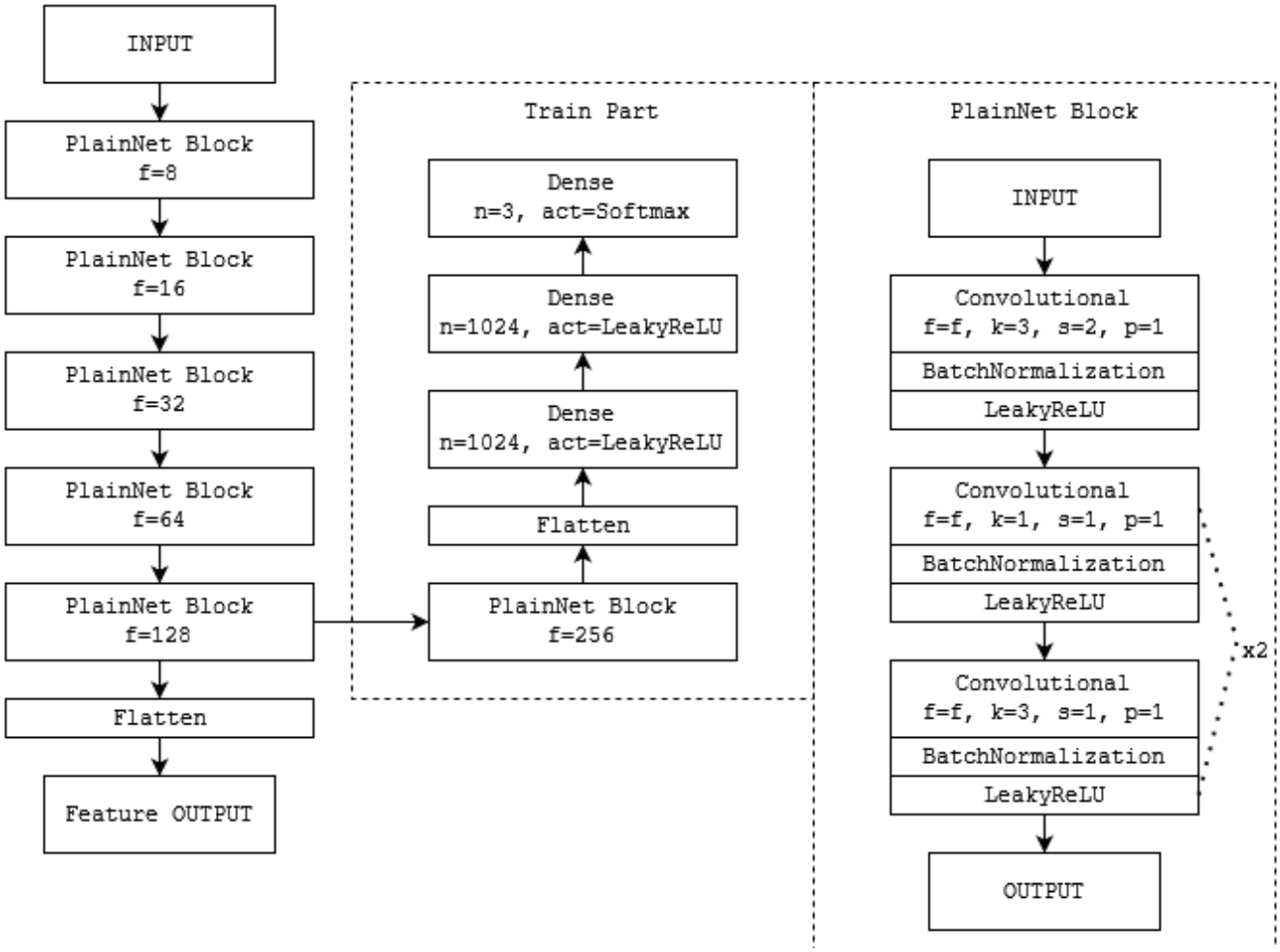



**Figure 4.** Diagram of CNN of ID Assigning System

counterpart. The schematic representation of this mode is depicted in Figure 3.

In contrast to the accuracy priority mode illustrated in Figure 3, the current mode is characterized by having just two outputs. Consequently, the output shape is simplified, maintaining the consistent output size criterion. The accuracy priority mode generates an output shape of (1, 7, 8400) based on a 640x640-pixel input size. Conversely, the speed priority mode, under the same input size criterion, significantly reduces the output array's coordinate count by over four times, resulting in a shape of (1, 7, 2000). This reduction not only streamlines processing, enhancing recognition speed, but also leads to reduced computational overhead.

Regarding the backbone structure, the current model exhibits a certain level of simplification compared to the accuracy priority mode depicted in Figure 3. Within this revised framework, a solitary CSPResidualBlock is substituted with a down-sampling convolutional layer. The head section is designed with increased simplicity to achieve the desired two outputs. While ResidualBlocks remain unused during upscaling, they are employed once during the downscaling process. This configuration facilitates the achievement of dual outputs.

*4. ID Assigning System for Object Detection*

TensorFlow Lite exclusively provides object recognition results in the form of coordinates and corresponding classes. However, it lacks a built-in mechanism for assigning identification (ID) tags to individual object features. Consequently, it becomes imperative to implement a custom system that can efficiently allocate IDs within TensorFlow. By doing so, we can effectively track and differentiate specific objects of interest. These unique IDs serve the purpose of distinguishing between various objects, such as differentiating between "Person 0" and "Person 1."

The ID allocation system comprises two distinct steps. Initially, feature extraction is performed using Convolutional Neural Networks (CNNs). Specifically, a CNN model called PlainNet, trained on datasets encompassing persons, dogs, and cats, is employed to extract features. Figure 4 visually illustrates the structure of this CNN model. In the following step, the system proceeds to extract features and subsequently evaluates the similarity of feature maps in order to either

reassign existing IDs or allocate new IDs. The following formula-based on cosine similarity formula is utilized for this purpose:

$$Similarity(F_{prev}, F_{curr}) = ReLU(\frac{\sum_{i=1}^{n} F_{prev_i} * F_{curr_i}}{\sqrt{\sum_{i=1}^{n} F_{prev_i}{}^2} * \sqrt{\sum_{i=1}^{n} F_{curr_i}{}^2}})$$

In accordance with this formula, in the event that a feature map exhibits a similarity of 50 percent or higher to an already existing feature map, it shall be reassigned the identical ID associated with said feature map. Conversely, if the similarity is below the designated threshold, a new ID shall be assigned and the corresponding feature map shall be associated with the newly assigned ID and subsequently stored.

*5. Object Tracking*

Experimental trials were conducted to track physical objects by connecting the application to the cradle device using the cradle's API. The API code controlling the cradle within the application was based on the code provided by Pivo. The cradle control functions included 'turnLeft,' 'turnRight,' 'turnLeftContinuous,' and 'turnRightContinuous.' To achieve smooth and continuous movement, the 'turnLeftContinuous' and 'turnRightContinuous' functions were employed, both of which continue rotating at the speed set before the function call if no speed value is provided as an argument. The rotational speed value represents the time it takes for the rotation to complete a full 360-degree cycle, rather than denoting angular velocity. To clarify, setting a rotational speed of 10 implies that the object will complete one full 360-degree rotation in a span of 10 seconds. This relationship can be expressed as follows:

$$rt = 360°/v_{actual} \quad (1)$$

In this equation, the variable 'rt' (Rotational Time) denotes the time required for a complete 360-degree rotation, serving as an input for the cradle control function. It functions as a time indicator for the cradle to complete a full rotation. Likewise, the variable $v_{actual}$ represents the cradle's real-time rotational speed, measured in angular velocity (degrees per second). Hence, taking into account the characteristic that 'rt' increases as the cradle's actual rotational speed decreases and decreases as it accelerates, the calculation of rotational time is facilitated.

The cradle's rotation speed is defined by the duration required to complete a full 360° rotation. Additionally, the system is designed to offer adjustable cradle rotation speed, enabling both rapid and gradual cradle rotations based on the object's proximity to the camera's central focal point. The extent of object displacement is determined by the object's x-axis coordinates relative to the camera's central point. Moreover, the cradle's rotation duration is intentionally configured to decrease as the object approaches the center of the camera's image. This feature is vital in preventing situations where, even if the object remains stationary, the cradle's failure to stop promptly could result in the object failing to align with the central point of the camera's image, resulting in lateral movement. Consequently, by capitalizing on this unique attribute of the cradle, the following formula is employed to calculate the rotation time, based on the distance between the object and the screen's central point.

$$MD = \begin{cases} 0 \ (if \ 0.4 < OP < 0.6) \\ (OP - 0.5) \times 2 \ (else) \end{cases} \quad (2)$$

$$RT = \begin{cases} \frac{6}{MD} \ (if \ MD \neq 0) \\ 0 \ \ (if \ MD = 0 \ ) \end{cases} \quad (3)$$

*OP = Object Position, MD = Moving Distance, RT = Rotational Time*

Equation 2 delineates the preprocessing steps involved in determining the object's x-axis central coordinates. In this equation, 'OP' (Object Position) signifies the x-axis coordinate of the object. It's crucial to emphasize that, for the sake of computational efficiency in this experiment, object positions have been standardized within a range of 0 to 1, measured in pixels. The distance from the screen center to the object is represented as 'MD' (Moving Distance). Specifically, when the object is aligned with the camera's center, it is designated as 'P=0.5'. If the object is positioned at the far right edge of the camera, it is expressed as 'P=1', while 'P=0' indicates the object is at the far left edge of the camera. It's worth mentioning that the cradle used in this study does not involve tilting, and therefore, vertical (y-axis) movement of the object is not considered.

In this equation, when the object's position (OP) falls within the range of 0.4 to 0.6, near the screen's center (OP=0.5), we designed it to return MD=0. Here, MD represents the distance from the screen center to the object, as previously explained, normalized values. Because the distance from the screen's central point, which is at 0.5, is minimal, it indicates that there is no need for the cradle to move, even if the object is outside this area. For positions outside this range, we subtract 0.5 from OP and then multiply the result by 2, ensuring it falls within the range of -1 and 1. When MD approaches 1, it signifies the object's position at the far-right edge. In such cases, the cradle should rotate more rapidly to the right for repositioning the object at the camera's center. Conversely, when MD approaches -1, it indicates the object's position at the far-left edge. In such instances, the cradle should move in the opposite direction to reposition the object located at the left edge back to the camera's center.

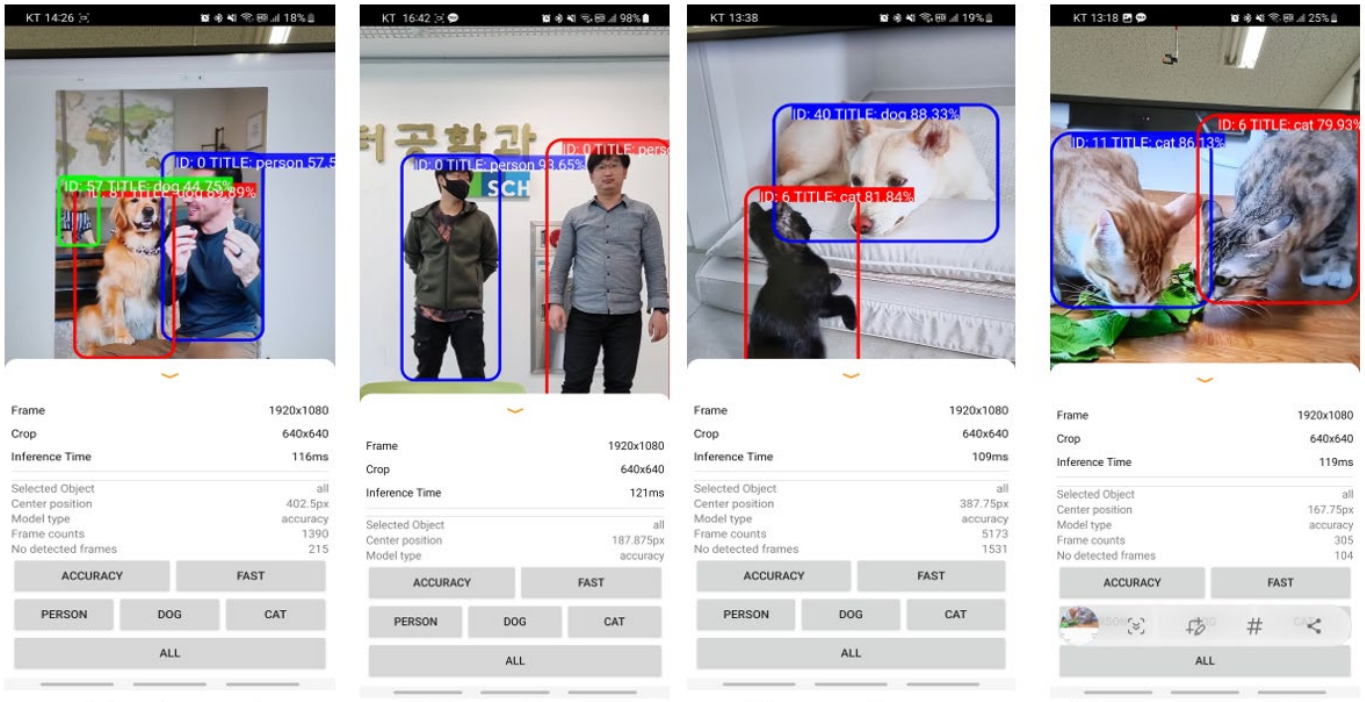
In Equation 3, RT (Rotational Time) is recalculated as the time required for the cradle to complete a 360° rotation based on the MD value. In the cradle control section, the API provided by the PIVO manufacturer was employed. The RT value spans a range from 6 to ∝. When the RT value is 6, the cradle rotates at its maximum speed, and as RT increases towards ∝, the cradle's rotation rate decreases towards zero value. The design ensures that as MD approaches 1 or -1, the

RT value tends towards 6. Conversely, when the object moves closer to the screen's center, causing MD to converge to 0, RT approaches ∝, resulting in a slower rotation speed. Furthermore, to prevent a ZeroDivisionError when the preprocessed MD value equals 0, RT is set to return 0 in such instances. Subsequently, the obtained RT values are applied to rotate the cradle. When RT is positive, the cradle rotates to the right, and when it's negative, it rotates to the left. If RT is 0, the cradle remains stationary, and no rotation occurs.

*6. Visual Display Implementation*

Figures 6 and 7 provide an overview of the object recognition component within our visual display application. The object recognition screen offers several key features. Firstly, the two models in the application utilize different input sizes. The accuracy priority mode operates with an input size of 640x640, while the speed priority mode uses an input size of 320x320. Secondly, users, particularly content creators, have the option to select between the accuracy priority mode and the speed priority mode. Thirdly, this screen incorporates an object selection function, enabling users to specify the objects they wish to recognize. Users can choose to detect people, dogs, cats, or all objects.

Furthermore, the screen displays an assigned ID above the recently recognized object. This ID is determined by analyzing the characteristics of the identified object, including its size, shape, and color. The assignment process ensures that the same object consistently receives the same ID, facilitating object tracking over time. In the application, the object selection and ID assignment features are important for the accurate and reliable recognition of objects. By allowing users to select the objects to be recognized, the application can be tailored to the specific needs of the user.




**(a)** 3 Classes of Smartphone with Accuracy Priority Mode

**(b)** People on Smartphone with Accuracy Priority Mode

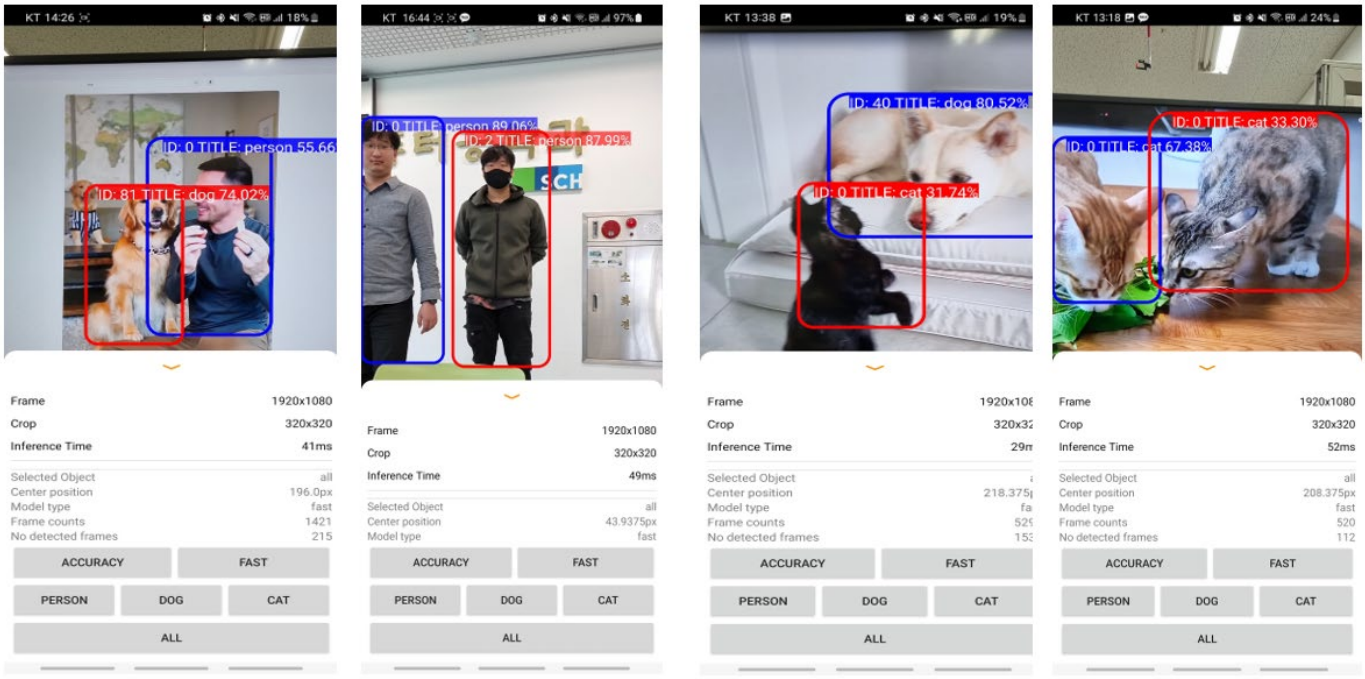
**(c)** Dog and Cat on Smartphone with Accuracy Priority Mode

**(d)** Cats on Smartphone with Accuracy Priority Mode

**Figure 5.** Examples of Screenshots with Accuracy Priority Mode

## III. RESULTS

In Figure 6, samples of research outcomes were depicted within the real-world experimental environment. This environment replicates a scenario where physical objects are captured by the camera, and subsequent experiments are conducted accordingly. It is important to note that no specific measures were taken to counteract the potential performance degradation resulting from the smartphone's heat buildup in this actual experiment. Additionally, the inclusion of an ID allocation system introduces additional computational resource consumption, which leads to a slightly reduced processing rate compared to the inherent capability of the model itself.




**(a)** 3 Classes of Smartphone with Speed Priority Mode

**(b)** People on Smartphone with Speed Priority Mode

**(c)** Dog and Cat Smartphone with Speed Priority Mode

**(d)** Cats on Smartphone with Speed Priority Mode

**Figure 6.** Examples of Screenshots with Speed Priority Mode

## IV. Conclusion

In conclusion, the increasing demand for real-time moving object detection systems in artificial intelligent smartphone cradles has prompted the development of two approaches in this paper. The accuracy priority mode and the speed priority mode were proposed to strike a balance between speed, high accuracy, and fast inference rate. Leveraging the cradle servo motion algorithm crafted for positioning mobile entities at the camera screen's center, the implementation of object tracking was achieved.

The accuracy priority mode leveraged the CSPNet structure integrated with ResNet based on the YOLOv8 architecture, achieving an impressive accuracy level of 93% based on the mean average precision at 50 intersections over union (mAP50) metric. It employed a shallower network depth and omitted bottleneck layers to enhance accuracy while maintaining a fast inference rate. This mode exhibits comparable accuracy to YOLOv8 while demanding fewer computing resources in smartphone environment.

On the other hand, the speed priority mode prioritized fast inference rates of 50 fps and achieved a commendable accuracy rate of 90% by reducing the number of layers. Among the various YOLO models, YOLOv5n is known as the shallowest architecture. The speed priority mode surpasses YOLOv5n in terms of computational resource requirements while maintaining a competitive level of accuracy. Notably, the accuracy of this is not significantly compromised compared to other models. The combination of accuracy reduces computational resource requirements, and high speed performance makes these models well-suited for detecting real-time moving multiple objects in one frame in a smartphone environment.

This study achieved successful real-time and accurate recognition of objects like humans, dogs, and cats. Nonetheless, the complexity of the CNN-based feature extractor in the ID assignment system introduced a notable drawback, leading to significant reductions in actual inference speed. Especially evident in the speed priority mode, the mode itself demonstrated an impressive inference speed of up to 50fps; however, object recognition led to a drop in frame rate to below 30fps. Consequently, there is a pressing need to simplify the feature extractor to minimize the loss of inference speed. Furthermore, certain cases unveiled scenarios where objects were assigned different IDs due to variations in perspectives, even when they were the same object or should share the same ID. Consequently, essential improvements like broadening the training classes of the feature extractor are required to address this concern.

Lastly, the practical implementation of tracking and controlling physical objects using a servo motor was achieved. This servo motor, part of a smartphone-compatible cradle, offers free rotation capabilities spanning 360 degrees. The cradle device, in conjunction with a dedicated API, facilitated the derivation of necessary formulas for object tracking and enabled us to conduct experiments with real objects. The results indicated the MAE value of 0.107 in the accuracy priority mode and 0.103 in the speed priority mode, significantly outperforming the YOLOv8s model with the MAE value of 0.223. This represents roughly a two-fold reduction in error compared to using the YOLOv8s model without our proposed

method. Hence, employing our approach can lead to enhanced tracking accuracy compared to conventional methods. However, in both the Accuracy Priority Mode and the Speed Priority Mode, the MAE values of the servo cradle were 0.107 and 0.103 on the average, representing a tracking distance of roughly twice as much in YOLOv8s.

## ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2023-2018-0-01799) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)

## REFERENCES

[1] Hwang, C.; Liao, G. Real-Time Pose Imitation by Mid-Size Humanoid Robot With Servo-Cradle-Head RGB-D Vision System, IEEE Transactions on Systems, Man, and Cybernetics: Systems 2019, 49, 181-191.

[2] Asali, M.O.; Saripudin; Indriyanto, T.; Trilaksono, B.R. Real-time Moving Object Video Tracking using Support Vector Machines for Visual Servo Application, In Proceedings of 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), Yogyakarta, Indonesia, (13-15 March 2019)

[3] Shu, Z.; Chao, F.; Nannan, C.; Shaosheng, L. The Design and Implementation of Automatic Camera Servo System Based on Improved Frame Differential Algorithm, In Proceedings of 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), Dengleng, (08-10 August 2011)

[4] Wang, C. Y., Liao, H. Y. M., Wu, Y. H., Chen, P. Y., Hsieh, J. W., & Yeh, I. H. (2020). CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 390-391).

[5] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[6] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

[7] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

[8] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[9] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13 (pp. 740-755). Springer International Publishing.

[10] Jocher, G., Chaurasia, A., & Qiu, J. (2023). YOLO by Ultralytics (Version 8.0.0) [Computer software]. https://github.com/ultralytics/ultralytics

[11] Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path aggregation network for instance segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8759-8768).

[12] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2117-2125).

[13] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 37(9), 1904-1916.

[14] Jocher, G. (2020). YOLOv5 by Ultralytics (Version 7.0) [Computer software]. https://doi.org/10.5281/zenodo.3908559

[15] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7464-7475).

# Multi-agent Reinforcement Learning for Stacking Meta-Model: An Application to Driver Emotion Classification

SeoHee Kim[1], Eunseo Jung[1], Hyojin Shin[1], and Jiyoung Woo[1*]
[1]*ICT Convergence Soonchunhyang University Asan, Republic of Korea*
[2] *University Asan, Republic of Korea*

*Contact: jywoo@sch.ac.kr

***Abstract*— Recently, there has been an increasing interest in the effects of stress and negative emotions on driving performance and road safety. The objective of this study is to develop a lightweight stacking model that classifies drivers' emotions into seven distinct categories by combining statistical electroencephalography data, psychological survey data, and driver behavior data. Using feature-based machine learning models instead of deep learning models can save memory and time. Therefore, our objective is to effectively combine individual machine learning models using reinforcement learning and achieve optimal performance by combining strong and weak learners. The final results show that the meta-model achieves the highest performance when using a decision tree, with an accuracy of 0.8543 and an F1-score of 0.8462.**

## I. INTRODUCTION

Driving is an important activity in our lives, but long journeys, traffic congestion, and rule violations can cause stress and negative emotions that impair driving performance and lead to dangerous situations [1].

Advances in science and technology have increased interest in research that collects and analyzes drivers' biometric signals to classify their emotions. Driver state monitoring is anticipated to become a mandatory requirement in automobiles, beginning with the European New Car Assessment Programme, further expanding the application of technologies to detect driver state and emotions. Currently, Kia Motors in South Korea is developing the real-time emotion adaptive driving (READ) system. The system includes technology that recognizes the driver's biometric signals and adjusts various sensory elements in the car in real time to optimize the vehicle's interior space according to the driver's emotional state and situation [2]. These developments can play a critical role in maintaining and improving driver safety.

Previous research on driver emotion classification has used various types of data, including images, voice, and biometric signals. While image data have been primarily used, they have the limitation of being affected by factors such as lighting and weather conditions. Voice data can be affected by the gender of the driver; therefore, special adaptations are needed to address this issue. Studies using biometric signals such as EEG, heart rate (HR), ECG, and exploratory data analysis (EDA) have classified drivers' emotions, but mostly into a maximum of five emotions [3-6].

This paper presents the development of an artificial intelligence model to classify drivers' emotions into seven distinct emotional states. The model integrates data closely related to the autonomic nervous system responsible for biological functions, specifically electroencephalography (EEG) data, psychological survey data obtained from drivers, and driver behavioral data.

## II. MODEL

In the proposed model, each model is considered an agent, and the performance of the entire model is compared based on the combination of these agents, where the increase or decrease in performance is defined as a reward. We train toward the optimal model combination by weighting rewards based on the performance of each agent's individual model. Models trained in this manner are expected to show better performance in terms of efficiency and effectiveness than the conventional approach, which simply uses all individual models.

Agents are created in a number equal to the number of individual models and defined as AGn. To find the combination in the stacking model, each agent has two actions for model selection: selecting a model is defined as 1, and not selecting is defined as 0. These are defined as At−1 for past cases and At for current cases. The reward is designed to progressively select better $A_t$ values through episode-by-episode learning. This is termed the F1 delta and reflects the difference between the pre-episode and post-episode, as expressed in Equation (1).

Past rewards are defined as $F1_{t-1}$, and current rewards as $F1_t$. Through this, agents adjust $A_t$ to maximize rewards. In addition, weights are used to create a balance between strong and weak learners. The weight $w_n$ for each $A_t$ of $AG_n$ is based on the F1-score of each individual model, as defined in Equation (2). For weak learners, i.e., models with

low F1-score values, a method is proposed that assigns additional weight to the reward granted in that episode when they are selected. This allows these models to receive a larger reward. Even if a strong learner changes their action in a given episode and contributes to performance improvement, it receives less reward than weak learners because the weight is given as 1 − F1-score. The final weight, $w_{final\ n}$, is incorporated considering the number of models receiving the reward, as shown in Equation (3). This is reflected proportionally. The models that receive the reward are those whose actions have changed compared to the previous time. Through this approach, rewards are adjusted considering the strengths and weaknesses of the models, thereby enabling harmony between strong and weak learners. Ultimately, $R_{final}$ is defined based on a comparison before and after each episode and weighting, as shown in Equation (4).

$$F1_{delta} = F1_t - F1_{t-1} \quad (1)$$

$$w_{n,t} = 1 - F1_{n,t} \quad (2)$$

$$w_{normalized\_n,t} = w_n / \sum_{n,t}(w_n, t)\left\{ \text{ if } n | A_t \neq A_{t-1} \right\} \quad (3)$$

$$R_{t,n} = F1_{delta} \times w_{normalized\_n,t} \quad (4)$$

Based on the aforementioned reward formula, the current state is updated based on the actions of each $AG_n$. The Q-value represents the expected reward that can be obtained when a specific action is selected in a given state. The agent tries to choose the action with the highest Q-value to maximize the reward. The update of the Q-value is defined by Equation (5). In the formula, $\alpha$ represents the learning rate, which adjusts the extent to which previous values and new rewards are reflected. $Q(a)$ represents the Q-value for a specific action, and Rt is the reward for the current action.

$$Q(a) = Q(a) + \alpha * (R_t - Q(a)) \quad (5)$$

## III. Result

In this study, the PPB-Emo [36] public dataset was used to classify 40 participants who took part in a driving experiment (31 males, and 9 females). The results of the predictions using each individual model are shown in Table VIII. The reward was modified in the reinforcement learning model to adjust the importance of weak learners through weights, which allows for a balanced selection of weak learners. As a result, the weak learners were appropriately combined to converge for better performance. Specifically, when the meta-model was DT, the optimal combination consisted of SGD, DT, NB, and vanilla NB, all of which were equally weighted, resulting in the highest accuracy.

TABLE I
Experiment Results

| Reward | Meta-Model | Accuracy | F1-score |
|---|---|---|---|
| X | DT | 0.5883 | 0.5587 |
| | SGD | 0.4983 | 0.4612 |
| | RF | 0.6549 | 0.6324 |
| Weighting | DT | 0.8543 | 0.8462 |
| | SGD | 0.7388 | 0.7316 |
| | RF | 0.8109 | 0.8092 |

These results show that weaker models, even if they underperform individually, can compensate for the shortcomings of the other models and act synergistically to improve predictive performance when combined. This study clearly showed that even when models with excellent performance are combined, improvement is not guaranteed, and it is not always possible to achieve the best result. This means that we can provide more robust and stable predictions by combining predictions from different models, signifying the need to consider interactions between models and complex data patterns.

TABLE II
Results of the Comparison with Other Models

| Model | EEG Data | Accuracy | F1-score |
|---|---|---|---|
| RF | Statistics | 0.8366 | 0.8288 |
| XGBoost | Statistics | 0.8345 | 0.8330 |
| DNN | Statistics | 0.6598 | 0.6498 |
| 1DCNN | Statistics | 0.7332 | 0.7069 |
| 1DCNN | Raw Data | 0.4857 | 0.4660 |
| RNN | Raw Data | 0.1512 | 0.1122 |

The study acknowledges that deep learning models, which are used across almost all fields, can be challenging to deploy on edge devices with limited computational resources, such as automobiles. In particular, the use of lightweight models is advantageous for accommodating data that are updated in real time. In this study, we sought to minimize the model size even when combining individual models by using a tree-based lightweight meta-model.

## IV. Conclusions

This study aimed to address the class imbalance problem and enhance the interpretability of MARL models for feature selection. In contrast to previous research, we introduced a reward distribution approach that assigns higher rewards to variables influencing the minor class. To achieve this, we utilized shap values to enhance the selection of variables impacting the minor class while also improving interpretability. Furthermore, through the combination of shap and MI, we crafted a comprehensive reward framework that achieved simultaneous performance enhancement and dimension reduction. This strategic approach allowed us to enhance the performance of the minor class, ultimately optimizing overall performance. However, this study did not consider the importance of variables in the medical context. Therefore, future research will focus on excluding or including medically

irrelevant variables to enhance the model and plan for additional investigations.

## REFERENCES


[1] Richard S Sutton and Andrew G Barto. Reinforcement learning: An introduction. MIT press, 2018.

[2] Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and Devon Hjelm. Mutual information neural estimation. In International conference on machine learning, pages 531–540. PMLR, 2018.

[3] Miron B Kursa and Witold R Rudnicki. Feature selection with the boruta package. Journal of statistical software, 36:1–13, 20

[4] [8] T. A. Gamage, L. P. Kalansooriya, and E.R.C. Sandamali, "An emotion classification model for driver emotion recognition using electroencephalography (EEG)," In 2022 Int. Res. Conf. Smart Comput. and Sys. Eng. (SCSE), vol. 5, pp. 76–82, 2022. doi: 10.1109/SCSE56529.2022.9905108

[5] [9] H. Halin, W. Khairunizam, W.A. Mustafa, M. A. Rahim, Z.M. Razlan, and S.A. Bakar, "Classification of human emotions using EEG signals in a simulated environment," In 2022 IEEE 13th Control and System Graduate Research Colloquium (ICSGRC), pp. 7–10, 2022. doi: 10.1109/ICSGRC55096.2022.9845131

[6] [10] C. Hieida, T. Yamamoto, T. Kubo, J. Yoshimoto, and K. Ikeda, "Negative emotion recognition using multimodal physiological signals for advanced driver assistance systems," Art. Life Robot., vol. 28, no. 2, pp. 388–393, 2023. doi: 10.1007/s10015-023-00858-y

[11] G. Du, Z. Wang, B. Gao, S. Mumtaz, K.M. Abualnaja, and C. Du, "A convolution bidirectional long short- term memory neural network for driver emotion recognition," IEEE Trans. Intell. Transport. Syst., vol. 22, no. 7, pp. 4570–4578, 2021. doi: 10.1109/TITS.2020.3007357