Article citation info:
Zalda K, Rulaningtyas R, Ain K, Rahma ON, Ittaqillah SI, Joseph AA, Azka RI, Fawza NN. Development of driver drowsiness classification based on EEG signals using long short-term memory method. Diagnostyka. 2026;27(2):2026215.
https://doi.org/10.29354/diag/225253.

DIAGNOSTYKA, 2026, Vol. 27, No. 2

e-ISSN 2449-5220
DOI: 10.29354/diag/225253

DEVELOPMENT OF DRIVER DROWSINESS CLASSIFICATION BASED ON EEG SIGNALS USING LONG SHORT-TERM MEMORY METHOD

Khouliya ZALDA [1], Riries RULANINGTYAS [1, 2], Khusnul AIN [1, 2], Osmalina Nur RAHMA [1, 2], Sayyidul Istighfar ITTAQILLAH [1], Annie Anak JOSEPH [3], Rafi Ihsana AZKA [1], Nabila Nayara FAWZA [1]

[1] Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia
[2] Biomedical Engineering Innovation Research, Universitas Airlangga, Surabaya, Indonesia
[3] Faculty of Engineering, Universiti Malaysia Sarawak, Malaysia
[*] Corresponding author, e-mail: riries-r@fst.unair.ac.id

Abstract

Driving accidents have a high prevalence cause of death with 1.3 million fatalities each year. Drowsy driving accounts for 26% of these accidents. A classification system for driver drowsiness was developed as a preventive countermeasure for the issue. The classification system was developed using Electroencephalograph (EEG) data recorded from four test subjects aged 19-24 during Virtual Reality (VR) driving simulations. The system categorizes drowsiness into four classes: awake, and three levels of drowsiness: mild, moderate, and severe, based on the reaction times measured during the Psychomotor Vigilance Test (PVT). EEG signals were processed using Discrete Wavelet Transform (DWT) to extract alpha, beta, and theta wave features. These features were then analyzed statistically using means and energy. This study compared the performance of four variations of Long Short-Term Memory (LSTM) models as a classification system. The most effective model was an LSTM with six hidden layers and 50 units per layer, achieving a test accuracy of 91% with the highest precision, sensitivity and F1-score, and accuracy average values of 0.93, 0.91, 0.92 across all classes.

Keywords: driver drowsiness; LSTM; EEG; virtual reality; discrete wavelet transform

List of Symbols/Acronyms

AUC – Area Under Curve
CSP – Common Spatial Pattern (if not used, you may remove)
db4 – Daubechies 4 Wavelet
DWT – Discrete Wavelet Transform
ECG – Electrocardiograph
EEG – Electroencephalograph
ELM – Extreme Learning Machine (if referenced)
EMG – Electromyograph
ESS – Epworth Sleepiness Scale
FFT – Fast Fourier Transform
FOV – Field of View
HRV – Heart Rate Variability (if referenced)
KSS – Karolinska Sleepiness Scale
LSTM – Long Short-Term Memory
NREM – Non-Rapid Eye Movement
OpenBCI – Open Brain–Computer Interface
PLI – Power Line Interference
PSD – Power Spectral Density
PVT – Psychomotor Vigilance Test
RNN – Recurrent Neural Network
SSS – Stanford Sleepiness Scale
VR – Virtual Reality
A_j[n] – Approximation coefficients at level j [μV]
b – Bias vector (subscript depends on gate type)
c_t – Cell state at time t
ĉ_t – Candidate cell state
D_j[n] – Detail coefficients at level j [μV]
E – Energy of sliding window [μV²]
f_t – Forget gate activation
g, k – High-pass (g) and low-pass (k) filter coefficients
h_t – Hidden state (LSTM output)
i – Sample index
i_t – Input gate activation
j – Wavelet decomposition level
μ – Mean value of the sliding window [μV]
n – Time shift
N – Number of samples in window
o_t – Output gate activation
ψ(t) – Mother wavelet (high-pass) function
ϕ(t) – Scaling (low-pass) function
W – Weight matrix (subscript indicates gate type)
x[n] – Discrete-time EEG signal [μV]

1. INTRODUCTION

Driving accidents have a high prevalence of causes of death, with a death rate reaching 1.3 million people each year worldwide. In Indonesia, the Central Statistics Agency (BPS) shows that in 2022 there were 139,258 accidents with a death rate

Received 2026-11-24; Accepted 2026-06-30; Available online 2026-06-30

of 28,131 people [1]. Drowsiness is one of the main factors causing accidents while driving, with 26.5% of driving accidents being caused by drowsiness. The risk of accidents will increase by 4-6 times when driving while drowsy compared to driving while conscious [2]. In a drowsy state, a person a person's alertness and cognitive responsiveness are significantly reduced. It is crucial to accurately assess the level of drowsiness to help drivers make informed decisions about whether to continue driving or take a rest. However, there is some challenge in doing research in drowsy driving due to lack of reliable measure of drowsy driving in Indonesia.

There are many methods being developed to classify the level of drowsiness while driving. These methods can be distinguished in two larger parts, which are self-assessment-based drowsiness classification and the objective measurement-based classification method. The self-assessment-based drowsiness classification involves using a standardized questionnaire that is handed on to the participant. Examples of it are Karolinska Sleepiness Scale (KSS), Epworth Sleepiness Scale (ESS), and Standford Sleepiness Scale (SSS). The validity of the questionnaire is shown by the correlation between the scale given by the participant with the behavioral sign of drowsiness such as driving slower and higher variations in the lane position [3]. However, this method can't track the drowsiness level continuously and potentially distract drivers if the frequency of the questionnaire is excessive.

The Objective measurement-based classification utilize the use of objective measurement such as driver reaction time or sensor placed on the vehicles or the driver [4]. This method is preferrable since it can track the driver or the vehicle condition continuously without distracting the driver. The objective measurement that has been used are Psychomotor Vigilance Test (PVT) [5], steering wheel angle [6], camera for lane edge detection [7], and Face detection [8]. The steering wheel angle method is not available in all cars. and the camera-based sensor is not reliable if the object of interest is obstructed [4]. Face detection mechanisms can fail due to abrupt lighting changes, such as during nighttime driving or when entering tunnels [9]. Similarly, lane edge detection may struggle when the lane markings are obstructed, faded, or missing, which are common challenges in classifying driver drowsiness[7]. Other potential objective measurement-based classification is by using physiological sensor. This is because the sensor can continuously monitor the driver's physiological conditions without intentional manipulation. There are several sensors that have been used, Electromyograph (EMG) [10], Electrocardiograph (ECG) [9], [11], and Electroencephalograph (EEG). From all of the physiological sensor, EEG become the best candidate since this sensor is also being used in polysomnography research because it can represent the changes in brain performance with high temporal resolution [5], [12], [13].

There are several research being done in the usage of EEG signal for driver drowsiness classification. [12] utilized EEG to classify drowsiness level in two class. Participants in the study were asked to perform a driving simulation on a tablet device under two conditions: drowsy and non-drowsy. During the simulation, EEG signals were recorded to capture brain activity corresponding to each condition. Then, the signal is being preprocessed using Discrete Wavelet Transform (DWT) and the Power Spectral Density (PSD) from each band is being calculated to be used as the input data for ELM classifier. Using this method, the researcher achieved a classification accuracy of 72.23%. [11] utilized Long Short-Term Memory (LSTM) as a Deep Learning classification method for the drowsiness classification. In this research, sensitivity of 81% and AUC of 0.88 was obtained using data collected from ECG signal. The author stated that the limitation of the study was the experimental environment was set on conventional driving simulator display. This could lead to the participant not feeling fully immersed in the driving activity and not seriously fighting against drowsiness during driving. However, the use of simulator remains essential, as real-world driving presents significant risks and high costs [14].

Based on the background, in this research, the use of LSTM and Virtual Reality (VR) in driver drowsiness classification task is being explored. VR display surrounds the user's Field of View (FOV), allowing the users to have depth perception. This caused the user to fully immersed and interact with the virtual environment compared to the conventional LCD screen which can only display object on a flat surface [15]. Validity of this display for driving simulation is proved by [16], which stated that there was no significant difference in average speed and FOV degree between the simulation and real world driving conditions. LSTM is a part of deep learning. Deep learning has an advantage over conventional machine learning as it is capable of learning from high dimensional data even when the data is partially corrupted by noise. This means minimal preprocessing required to clean the data. LSTM is known for its ability to retain memory over time. This is due to its gating mechanisms that regulate the flow of information through the network [17]. This enables the model to effectively recognize patterns in sequential data, making it particularly suited for tasks such as EEG signal classification in driving drowsiness condition. The EEG signal processing, including feature extraction and LSTM model training, was executed digitally using Python. The processing pipeline utilized standard open-source libraries, primarily SciPy for signal filtering and TensorFlow/Keras for building the deep learning architecture. Given the highly efficient and lightweight nature of the extracted statistical features, the entire

computational process did not require High-Performance Computing (HPC) or specialized graphic processing units (GPUs). The training was effectively and seamlessly handled using a standard local desktop environment, ensuring high reproducibility for future practical implementations. The information obtained from the preprocessing result will be the input for LSTM model to classify the drowsiness level into 4 classes, which are normal, light drowsy, intermediate drowsy, and high drowsy. The level of drowsiness corresponded to the PVT test result, as the PVT is a reliable method to assess drowsiness-based reaction time which reflects the driver's performance [18]. The main contributions of the study are summarized as follows: i) developing a method for driver drowsiness level classification based on LSTM model ii) improving the data collection procedure by incorporating VR display to provide realism in the experimental environment. Hopefully, the driver drowsiness classification method can be a way to reduce the number of accidents caused by drowsiness by giving accurate classification results to help drivers make informed decisions about whether to continue driving or take a rest during driving session.

2. MATERIALS AND METHODS

2.1. Data collection from participants

The data used in this study was EEG data collected from 4 male participants in age range of 19-24 years without history of neurological disease. The gender selection was done to ensure proper contact between the EEG cap and the scalp to minimize the noise caused by poorly attached electrodes. The data collection took place in 14.00-17.00 in which most drowsy related accident occurred due to body's circadian rhythm [2], [3], [19]. The participants were told to have proper sleep and not allowed to drink caffeinated drinks such as coffee, or energy drinks in a range of 8 hours before the experiment began. This was to minimize the effect of improper sleep or caffeine related effects to the experiment, as both can affect the driver's performance [20]. Detailed information regarding the experiment were given to the participants. The study complies with all regulations and informed consent was obtained from all participants.

In this study, the participants were asked to drive on a driving simulation application (City Car Driving, Multisoft, Russia). Before the data collection began, the participants first familiarized themselves with the driving simulator. After that, an EEG cap was worn by the participants. The EEG device used in this research was an 8 channel OpenBCI EEG cap with 250 Hz sampling frequency. The cap was used to ensure that the channel position followed the standardized 10-20 system. Conductive gel then was applied to each electrode to further increase the conductivity. Logitech G29 force-feedback steering wheel with three-foot pedals was used in the experiment as additional hardware to add realism to the driving simulation. To further add realism, a Meta Quest 2 VR was used to project the driving simulation display. This created the sense of presence to the participant by surrounding the participant's Field of View (FOV) with the virtual driving environment. The device setup was in the participant is shown on Fig. 1.

Fig. 1. The Electroencephalograph (EEG), Virtual Reality (VR) display, and the steering wheel setup on the participant during data collection

The participants were instructed to drive for 2 hours that divided into 8 sessions which session last for 15 minutes in free driving scenario with maximum speed of 60km/h. The simulation was set on afternoon at a freeway and low-density town with 10% traffic to simulate monotonous environments which could trigger driving drowsiness [2]. Between each session, the participants were instructed to pause the driving activity and did the PVT test to test their reaction time. In addition, the participants subjective scale was also collected using Karolinska Sleepiness Scale (KSS) questionnaire. The experimental procedure details are shown in Fig. 2. Recorded EEG signal then classified based on their PVT test results as showed on Table 1. The use of PVT test as the ground truth was backed up by the fact that PVT test has intra and intersubject variability, so that the result can represent the changes in each subject objectively [5].

Table 1. Drowsiness level classification based on reaction time from PVT test [21]

Reaction Time (ms)	Drowsiness Level
>250	Normal
250-350	Light Drowsiness
350-600	Medium Drowsiness
>600	Heavy Drowsiness

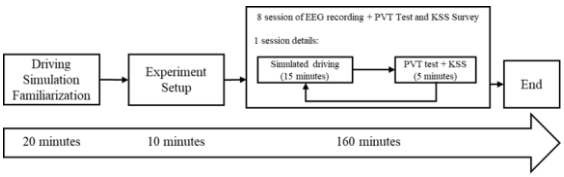

Fig. 2. Data Collection Timeline

2.2. Data preprocessing

The EEG signal collected from the experiment was preprocessed to remove noise before the feature extraction, as the presence of noise can degrade signal quality and affect the classification result. The noise in EEG signal can come from physiological noise such as muscle activity from eye and jaw movement, or from environmental factors such as powerline interference (PLI), decreased electrode conductivity over time, and movement artifact. While the noise is not often directly visible in the time domain, it can be localized effectively in the frequency domain where certain frequency bands correspond to common noise source, such as PLI in the frequency 50/60 Hz [18]. The removal of the noise can be carried out using various filtering algorithms. In this research filtering process was carried out using 4th order butterworth bandpass filter at frequency 4-64 Hz and notch filter at 50 Hz because the useless physiological and environmental noise occurred outside the frequency range [19].

EEG signals can be decomposed into distinct frequency bands. The frequency band used in this study are theta (4-8 Hz), alpha, (8-13 Hz), and beta (13-30 Hz). Each band is associated with a specific neural process associated with drowsiness conditions. Beta band is commonly found in wakefulness indicating cognitive process. Alpha band also found in awake condition, but in suppressed state. The activity will be increased in idle and relaxed state such as when a person is going to sleep. The theta wave only occurs when a person enters Non-Rapid Eye Movement (NREM) sleep stage. During this stage, the alpha and beta band activity is weakened [18]. The presence of heightened alpha or theta when a person engages in an activity such as driving is dangerous because it might indicate alertness reduction of the driver to external stimulus.

2.3. Discrete wavelet transform (DWT)

To decompose EEG signals into separate frequency band, Discrete wavelet Transform (DWT) was applied to the filtered signals using Daubechies 4 (db4) as the mother wavelet its ability to balance time-frequency localization with the morphological characteristics of neural oscillations. Unlike simpler wavelets like Haar**,** db4 provides a higher number of vanishing moments, which enhances its capability to filter out noise while preserving the essential physiological features required for high-accuracy classification. This specific choice is crucial because db4 provides an optimal number of vanishing moments that significantly enhances the extraction of morphological neural oscillations. This rationale is highly supported by wavelet thresholding methodologies in EEG processing, demonstrating that other simpler wavelets are less effective for this specific physiological filtering [6]. The DWT equation can be described as follows:

For a discrete signal x[n], DWT is computed by doing convolution to the signal with scaled and shifted version of a wavelet function ψ(t) called mother wavelet and a scaling function of ϕ(t) which resulted in a set of coefficients representing the signal in different resolutions [22].

Scaling Function (Approximation coefficients):

$$A_j[n] = \sum_k x[k].\phi_{j,n}[k] = \sum_k x[k].2^{-\frac{j}{2}}\phi\left(2^{-j}k - n\right) \quad (1)$$

In equation 1, $A_j[n]$ are the approximation coefficient at scale j and position n and function to capture the low frequency components of the signal.

Wavelet Function (Detail coefficients):

$$D_j[n] = \sum_k x[k].\psi_{j,n}[k] = \sum_k x[k].2^{-\frac{j}{2}}\psi\left(2^{-j}k - n\right) \quad (2)$$

In equation 2, $D_j[n]$ are the detail coefficient at scale j and position n and function to capture the low frequency components of the signal. For both equations, j is the scale (level) of the resolution, and n is the translation (shift in time). The signal is passed through both filters, with scaling function ϕ(t) act as a low pass filter and wavelet function ψ(t) as a high pass filter. The filtering process results in a downsampled signal at each level. The multiresolution of the wavelet properties is captured during this process, with each level focusing on finer details.

For the first level of the decomposition, the approximation and detail coefficient can be computed directly from the signals, resulting in simplified version of the equation showed at equation 3 and 4.

(Level 1 Approximation coefficients):

$$A_1[n] = \sum_k x[k].\phi_{1,n}[k] = \sum_k x[k].h[k-2n] \quad (3)$$

For the first level of the decomposition, the approximation and detail coefficient can be computed directly from the signals, resulting in simplified version of the equation showed at equation 3 and 4.

(Level 1 Detail coefficients):

$$D_1[n] = \sum_k x[k].\psi_{1,n}[k] = \sum_k x[k].g[k-2n] \quad (4)$$

With g and k represent the high pass and low pass filter coefficient derived from the wavelet function. For the next level, the equation can be generalized as shown in equation 5 and 6, with l as the decomposition levels:

Level *l* Approximation coefficients):

$$A_l[n] = \sum_k A_l[k].h[k-2n] \quad (5)$$

Level *l* Detail coefficients:

$$D_l[n] = \sum_k D_l[k].g[k-2n] \quad (6)$$

In this research, the DWT process was applied to the filtered signal at 4 level of decomposition. This was done to extract the alpha, beta, and theta frequency band from the signals. The decomposition process explained by Fig. 3.

Frequency band extraction using DWT resulted in the decomposition of the signal into multiple frequency bands for each electrode. For every frequency band in each electrode, the feature is further extracted using statistical methods. The statistical features were mean and energy. These two features were exclusively selected as they provide a

computationally efficient representation of the signals' spatial and spectral power without overloading the network. Although potential correlation (multicollinearity) may exist between them, no additional features were needed because the gating mechanisms of the subsequent LSTM architecture independently regulate the influence of each input, inherently tolerating collinearity to extract meaningful temporal patterns [8] [9]. The feature extraction process was carried out by applying sliding windows, which apply the statistical operation over a predefined window length then do the overlapping process by shifting the windows into the next data point as illustrated in Fig. 4. In this research, the windows length is 5 second. This specific duration was justified because it is long enough to adequately capture complete cycles of low-frequency neural oscillations, particularly the theta band (4-8 Hz) which is critical for identifying NREM sleep transitions [9]. Simultaneously, it is short enough to maintain the high temporal resolution needed to detect rapid neurophysiological shifts associated with brief lapses in driver vigilance [10].

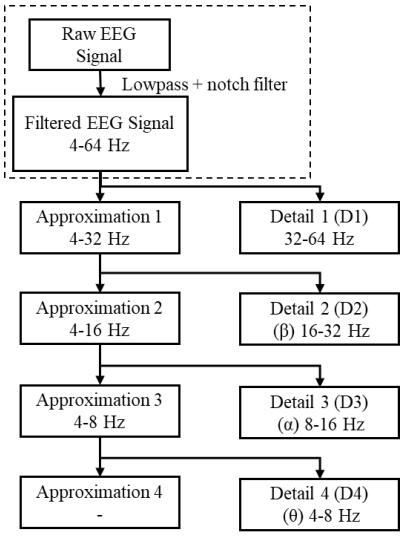

Fig. 3 EEG Frequency Band Extraction using 4 level of DWT Diagram

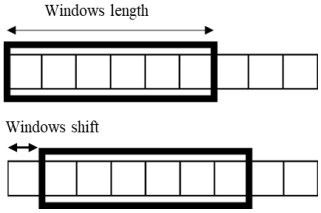

Fig. 4 Sliding Windows Illustration Used for Statistical Feature Extraction

The equation used to calculate the statistical feature in the sliding windows can be described as follows, with x[i] as the signals obtained from DWT process and N is the length of the windows.

Mean (μ):

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x[i] \tag{7}$$

Energy (E):

$$E = \sum_{i=1}^{N} x[i]^2 \tag{8}$$

2.4. Long short term memory (LSTM)

This research used LSTM as the method for the driver drowsiness classification in VR simulated environment. LSTM was first introduced as a solution to overcome vanishing gradient problem while training model in conventional Recurrent Neural Networks (RNN). LSTM is known for its ability to retain memory over time. This is due to its gating mechanisms that regulate the flow of information through the network. This made LSTM excels for sequence related task such as text generation, time series forecasting and anomaly detection[17]. The LSTM memory cell structure is visualized at Fig 5 and can be explained as follows

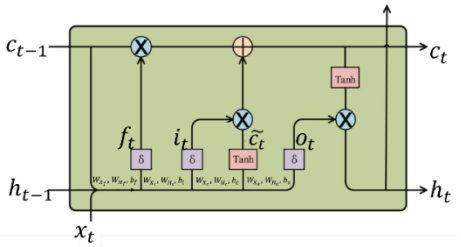

Fig. 5. Visualization of LSTM Memory cell calculation

At forget gate (f_t), sigmoid function was applied to the weighted sum of the current input (x_t) and the previous hidden state (h_{t-1}). The W_{X_f} is the weight matrix that corresponds to the current input (x_t), capturing how much influence the current input has on the forget gate's decision. W_{H_f} is the weight matrix associated with the previous hidden state (h_{t-1}) reflecting how much of past information impacts the current forget decision. The b_f is the bias in the forget gate to enable independent adjustment of the gate's output.

$$f_t = \sigma(W_{X_f}x_t + W_{H_f}h_{t-1} + b_f) \tag{9}$$

Same logic can also be applied for the input gate (i_t), with the difference in weight matrices used in the input gate are (W_{X_i}) and (W_{H_i}) This allows the input gate to determine how much of the new information should be allowed into the cell state based on both the current input and the past hidden state.

$$i_t = \sigma\left(W_{X_i}x_t + W_{H_i}h_{t-1} + b_i\right) \tag{10}$$

At the output gate (o_t), the W_{X_0} and W_{H_0} determine how much current and past information influence the output gate result.

$$o_t = \sigma\left(W_{X_0}x_t + W_{H_0}h_{t-1} + b_0\right) \tag{11}$$

Cell state candidate ($\tilde{c}_t$) represent how much the potential information can be added to the cell state. Unlike other gate, the cell state candidate calculation is done by applying tanh function to the weighted sum of the current input (x_t) and the previous hidden state (h_{t-1}). In equation 12, the $W_{X_c}$and the W_{H_c}represent the weight matrix in the cell state. The use of tanh instead of sigmoid is to represent the potential information strength and direction.

Table 2. KSS dan PVT result in each participant

	Session	KSS (1-10)	PVT (ms)	Drowsiness Classification		Session	KSS (1-10)	PVT (ms)	Drowsiness Classification
Participant 1	1	5	363.54	Medium	Participant 3	1	3	194.67	Normal
	2	7	526.43	Medium		2	6	263.73	Light
	3	8	651.27	High		3	5	215.99	Normal
	4	4	324.25	Light		4	6	220.7	Light
	5	4	220.67	Normal		5	3	189.76	Normal
	6	5	345.4	Light		6	5	264.19	Normal
	7	7	496.66	Medium		7	7	367.24	Medium
	8	7	485.57	Medium		8	5	278.16	Medium
	Session	KSS (1-10)	PVT (ms)	Drowsiness Classification		Session	KSS (1-10)	PVT (ms)	Drowsiness Classification
Participant 2	1	3	220.67	Normal	Participant 4	1	4	178.22	Normal
	2	4	284.72	Light		2	3	181.26	Normal
	3	5	312.82	Light		3	5	310.22	Light
	4	6	488.37	Medium		4	5	262.81	Light
	5	4	345.4	Light		5	3	237.71	Normal
	6	6	485.67	Medium		6	5	269.47	Light
	7	7	633.79	High		7	6	382.63	Medium
	8	5	496.66	Medium		8	2	224.21	Normal

$$\tilde{c_t} = \tan\mathrm{h}\left(W_{x_c}x_t + W_{H_c}h_{t-1} + b_c\right) \quad (12)$$

The cell state update (c_t) in equation 13 represent the updated cell state at the current time step which will be passed to the next time step. The update is calculated by scaling the previous cell state (c_{t-1}) with forget gate output (f_t) to determine how much past information to keep. The new information represented by the candidate cell state $(\tilde{c_t})$ is scaled by the input gate output(i_t)

$$c_t = f_t c_{t-1} + i_t \tilde{c_t} \quad (13)$$

The output of the LSTM at current timestep is represented by hidden state (h_t). The output from the hidden state is used for the calculation for the next layer as (x_t) or in the next timesteps as (h_{t-1}). The output gate is calculated by multiplying the output gate (o_t), which controls how much of the updated cell state (c_t) should be output with the tanh result from cell state (c_t)

$$h_t = o_t \tan\mathrm{h}(c_t) \quad (14)$$

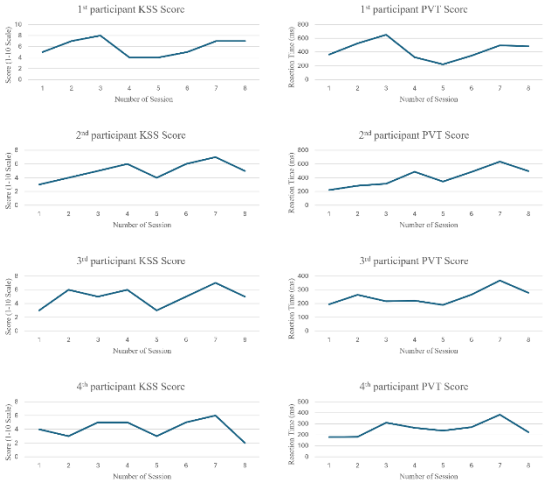

Fig. 6. KSS and PVT plot from Table2

Based on the 120 minutes of total EEG recordings across the four subjects, the 5-second non-overlapping window strategy yielded exactly 1,440 windows per participant, totaling 5,760 samples [3][4]. Considering 8 EEG channels, 3 frequency bands, and 2 statistical features, each sample successfully formed a 48-dimensional input vector (8 × 3 × 2 = 48 data points per window).

Based on the session-level labels reported in Table 2, the aggregate distribution across all four participants and eight sessions was as follows: Normal (13 sessions), Light (10 sessions), Medium (9 sessions), and Heavy (4 sessions), yielding an approximate ratio of 36%:28%:25%:11%. Applying a 5-second non-overlapping window to each 15-minute session and proportionally distributing the total of 5,760 samples reported in Section 2.4, the estimated number of samples per class was: Normal = 2,080 samples (36%), Light = 1,600 samples (28%), Medium = 1,440 samples (25%), and Heavy = 640 samples (11%). Following an 80/20 train-test split, the training set contained approximately 4,608 samples and the test set approximately 1,152 samples. While a moderate imbalance exists toward the Heavy class as the minority, the reliability of the classification results is substantiated by the per-class performance metrics in Table 3, where the best-performing model achieved average precision, sensitivity, and F1-score of 0.93, 0.91, and 0.92 across all classes. In this research, different number of LSTM layer and number of LSTM unit in each layer is being explored. There are 4 variations of model, which are 3 hidden layers with 10 LSTM unit in each layer, 3 hidden layers with 50 LSTM unit in each layer, 6 zhidden layers with 10 LSTM unit in each layer, and 6 hidden layers with 50 LSTM unit in each layer. In this study, the data were split into training, validation, and testing data with 80% data used as training and 20% were split into validation and testing. The data was then evaluated using the test data to create a confusion matrix for all class. From the confusion matrix, the evaluation metrics can then be calculated in each class using equation 15 –equation 18.

$$Accuracy = \frac{\sum_{i=1}^{C} TP_i}{\sum_{i=1}^{C}(TP_i + FN_i)} \quad (15)$$

$$Precision = \frac{TP_i}{TP_i+FN_i} \quad (16)$$

$$Sensitivity = \frac{TP_i}{TP_i+FP_i} \quad (17)$$

$$F1\ Score = 2x\frac{Precision_i\ x\ Sensitivity_i}{Precision_i + Sensitivity_i} \quad (18)$$

3. RESULT AND DISCUSSION

3.1. Data collection from participant

Raw Data was collected using the procedure explained in section 2.2. The participants were asked to drive in a simulated environment while EEG cap was attached. This allowed the researchers to observe and record the EEG signals produced by the participant during the driving session. Between each session, the subjective response and the reaction time of each participant was also recorded using KSS and PVT test respectively. The PVT test device was based on audio stimulus, meaning that the participant had to respond to the beeping sound by pushing a button. The sound produced by the device was given in random intervals so that the test was independent of aptitude and learning (inter and intra subject variability). The KSS and PVT test result was displayed on Table 2 and plotted on Figure 6. Figure 6 showed that the performance changes in reaction times are associated with the changes of perceived sleepiness in KSS. As detailed in Table 2, an observable proportional trend demonstrates that an increase in KSS scores consistently corresponds to longer PVT reaction times across all participants. This direct association aligns with existing literature, confirming that elevated subjective sleepiness directly impairs objective cognitive reaction times. While a formal Pearson or Spearman correlation was not computed, the within-subject monotonic trend observed across all four participants, where increasing KSS scores consistently corresponded to prolonged PVT reaction times, constitutes a robust qualitative validation, consistent with prior literature [4][5]. Given the small sample size (n=4), a single aggregate correlation coefficient would carry limited statistical power and risk masking individual variability, which is precisely what Table 2 and Figure 6 are designed to reveal. This result supported the study that was done by [3] which stated that KSS score had correlation with the driver performance. The driver performance was represented by the PVT result that align with the statement from [2] that stated driver drowsiness can cause attentional time lapses and slowed down reaction time.

The KSS and PVT scores in each participant during each session differed with each other. This was caused by the different situations the participants might experience during sessions. It increased with each session when the participants found the road condition was monotonous, such as on the highway or while stopping at the traffic light. On the other hand, the score decreased when sudden conditions occurred, such as hitting the map edge or bumping into other cars. This align with [9] that stated under stimulation during driving triggered the driver to enter the state of relaxation and lack of alertness which accumulated driver drowsiness.

During driving in simulation, the EEG of the participant was also recorded. The recording process was done using software provided by OpenBCI. The file was saved in the .txt format to be processed later in python. The first process is to plot the raw EEG signal collected from 8 electrode. The electrode was placed on FP1, FP2, C3, C4, T5, T6, O1, and O2. The example of the recorded raw EEG signal from one participant during 1 session of recording is plotted on Fig 7. It can be observed that all channels had different baselines, and the baseline experienced a drift over time. This was caused by the change of impedance value due to the electrode drying or shifting position from the original point when the experiment was started. Another noise that corrupted the EEG signal was PLI noise. The noise was visibly seen by zoomed in the signal as shown in Fig 8. To further confirm the noise, FFT transformation was applied to all signals, with example in channel O1 as shown in Fig 9. The unusually high magnitude was found in the frequency under 4 Hz and in 50 Hz. Figure 9 presents the normalized single-sided amplitude spectrum of raw EEG channel O1, where the y-axis represents amplitude in µV, obtained by dividing the raw FFT coefficients by N/2 (N=1,250, corresponding to 5 seconds at 250 Hz sampling rate). The prominent peaks below 4 Hz indicate baseline drift, and the peak at 50 Hz indicates Power Line Interference (PLI). Both noise sources were subsequently removed using a 4–64 Hz bandpass filter and a 50 Hz notch filter. PLI noise was caused by the use of electronic devices during the experiment, specifically the computer used to run the driving simulator and transmit the display to the VR device.

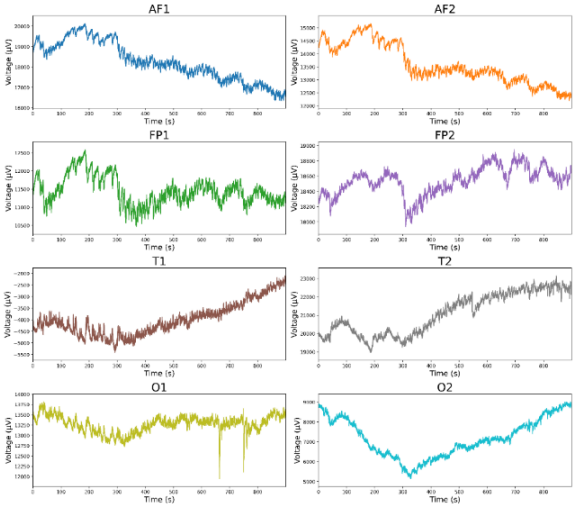

Fig. 7. Raw EEG signal plot from one participant during 1 session of recording (15 minutes), with: a) channel FP1 b) channel FP2, c) channel C3, d) channel C4, e) channel T5, f) channel T6, g) channel O1, h) channel O2. x-axis is the time in second and y-axis is voltage in microvolt

PLI noise was caused by the use of electronic devices during experiments, which were a computer used to simulate the driving simulator and send the display to the VR device.

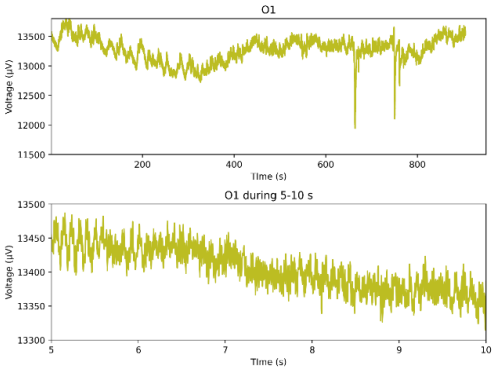

Fig. 8. Plot of channel O1, with: a) channel plot in 15 minutes recording b) zoomed in channel plot during 5 seconds recordings

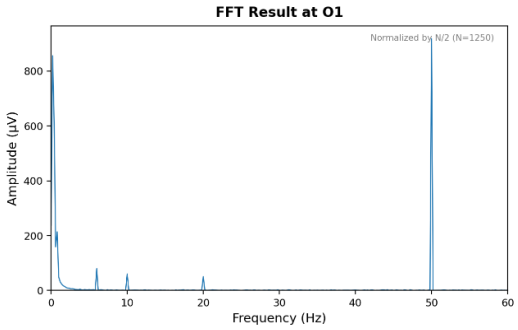

Fig. 9. Plot of FFT from raw data channel O1

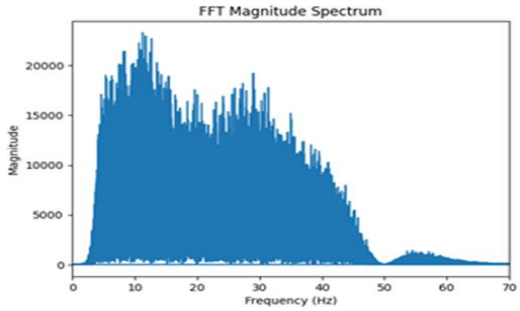

Fig. 10. Plot of FFT from filtered data channel

3.2. Preprocessing and Feature Extraction

The presence of noise can degrade signal quality and affect the classification result. The noise was then removed using 4th order butterworth filter in range of 4-64 Hz and notch filter at 50 Hz. This was done to remove the noise and the frequency outside the frequency of interest. The filtered result now had normalized baselines and contained frequency only in range of 4-64 Hz as shown in Figure 10 and Figure 11. The signal was then decomposed by applying DWT from equation 1 to equation 6. By using DWT, the signal could be analyzed in multiresolution without huge amounts of lost information. DWT transformation produced the approximation coefficient and detail coefficient, with detail coefficient D2, D3 and D4 represented three different frequencies band of EEG, beta, theta and alpha following the diagram on Figure 3 and plotted in Figure 10.

3.3. LSTM classification

In this research, LSTM model was used to classify the drowsiness level during driving. The classification steps consist of training and testing process. The data was labelled into 4 classes, which were normal, light drowsy, intermediate drowsy, and high drowsy. The data was then divided into training data and testing data with 80% of the data as training data. To be able to enter the classification process, the data first was transformed into a 3D array which consist of batch size, timesteps, and feature. The batch size allows the model to learn from multiple sequences, while the timesteps represent the number of the sequential input. The transformed data then enter the LSTM cell following the flow diagram illustrated in Figure 5, with each gate applied the equation from equation 9 to equation 14.

Four variations of model were used in this research with the difference on the LSTM layer depth and number of LSTM unit in each layer. The variations were 3 hidden layers with 10 LSTM unit in each layer, 3 hidden layers with 50 LSTM unit in each layer, 6 hidden layers with 10 LSTM unit in each layer, and 6 hidden layers with 50 LSTM unit in each layer. The data was trained using fixed hyperparameter for each model with 100 epochs of training.

The number of LSTM units determines the model's ability to capture patterns within the data. Larger unit counts in each layer allow the model to store and process more information at each time step, which is beneficial for learning complex patterns in long sequences. Numbers of LSTM layers impact the model's ability to learn hierarchical representations of sequential data, enabling the model to capture both short-term and long-term dependencies effectively to generalize across different datasets. Either too few layer or LSTM units can lead to underfitting, a condition where the model had too few parameters causing the model fails to capture essential patterns in the data. This showed in Figure 13. Where both training and validation accuracy

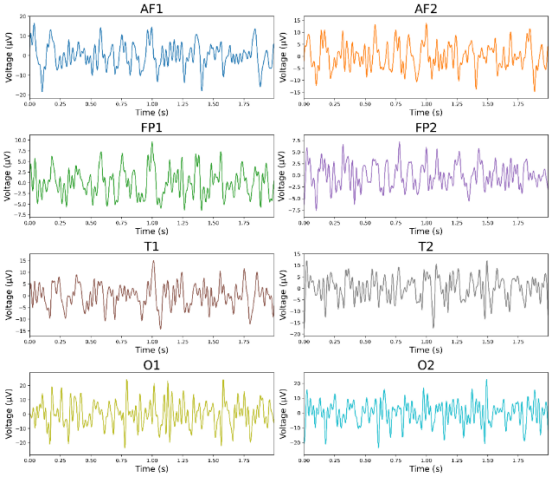

Fig. 11 Sample of Filtered EEG Signal Plot from One Participant During 2 Second of Recording

didn't improve much and showed little improvement over time. The condition is improved when the complexity of the model increased by using a higher number of layers and LSTM unit. The configuration of 6 hidden layers with 50 LSTM unit in each layer produced an increasing trend of the accuracy with small fluctuation after epoch 30, indicating a welltrained model as shown in Figure 14.

The model performances were then evaluated on test data by creating a confusion matrix so that precision, sensitivity, F1-score and accuracy can be calculated by applying equation 15 to equation 18. The results of the performance calculation were summarized on Table 3. The table further inferred that the model configuration affects the performance. From all the configuration, the configuration of 6 hidden layers with 50 LSTM unit

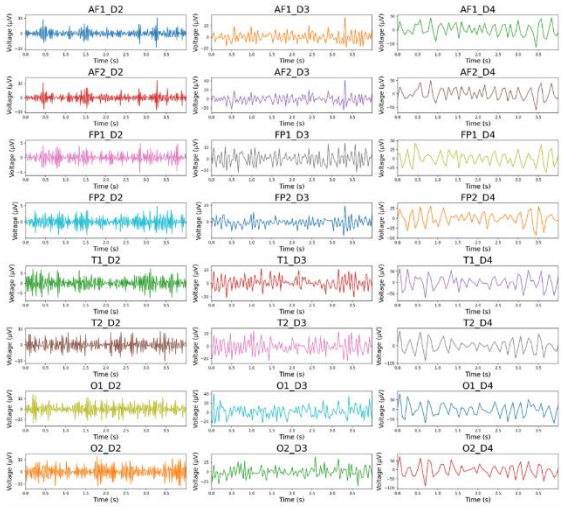

Fig. 12 Sample of DWT Result of EEG Signal Plot from One Participant During 5 Second Recordings

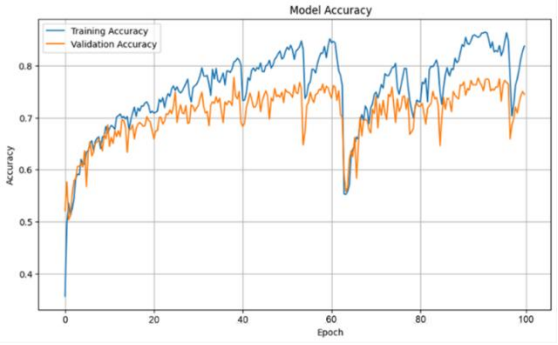

Fig. 13 – Model Training Accuracy Throughout Epoch During Training Process on 3 Hidden Layer with 10 LSTM Unit in Each Layer

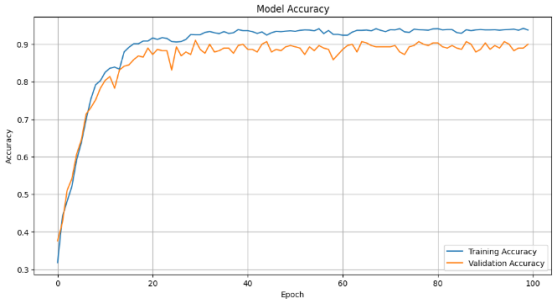

Fig. 14 – Model Training Accuracy Throughout Epoch During Training Process on 6 Hidden Layer with 50 LSTM Unit in Each Layer

in each layer achieved the highest precision, sensitivity, F1-score, and accuracy with average values of 0.93, 0.91, 0.92 and 0.91 across all classes. The balance precision and sensitivity in this configuration is preferred, as high level of sensitivity increased driver's confidence in the warning system by minimizing false alarm [4].

Table 3 Model Classification Accuracy in Each Configuration

6 Hidden Layer	10 LSTM Unit	Normal	0.81	0.79	0.8	0.81
		Light	0.73	0.87	0.79	
		Intermediate	0.99	0.96	0.97	
		High	0	0	0	
	50 LSTM Unit	Normal	0.9	0.72	0.8	0.91
		Light	0.84	0.95	0.89	
		Intermediate	0.99	0.99	0.99	
		High	1	0.97	0.99	
Configuration		**Drowsiness Level**	**Precision**	**Sensitivity**	**F1-Score**	**Testing Accuracy**
3 Hidden Layer	10 LSTM Unit	Normal	0.77	0.73	0.75	0.65
		Light	0.59	0.86	0.7	
		Intermediate	0.74	0.36	0.48	
		High	0.41	0.43	0.42	
	50 LSTM Unit	Normal	0.72	0.71	0.72	0.74
		Light	0.73	0.75	0.74	
		Intermediate	0.82	0.84	0.83	
		High	0.54	0.43	0.48	

Compared to the previous research done by [11], [12], this research showed higher performance in classifying driver drowsiness model although more class was being used. The high performance of the model can be attributed to the different use of the modality used to simulate the driving environment. Deep Learning classification such as LSTM is highly dependent on the data. The use of VR in this research allowed the participants to drive in an immersive environment that resembles real world conditions closer compared to the use of conventional screen which support [15]. This approach resulted in data that might resemble actual driving conditions so that it improved the model's performance.

4. CONCLUSION

The EEG signal-based driving fatigue classification system using the LSTM method in VR simulation can be created by creating an LSTM model by varying the number of hidden layers used and the number of LSTM units in each layer. The model is then trained using training data sourced from the EEG signal features of the participant when simulating driving on a VR device. The features used as LSTM input are the statistical features of mean, and energy extracted every 5 seconds. Each statistical feature is obtained from the wavelet transformation process of the EEG signal which produces alpha, beta, and theta waves.

The variation of the model created has validity for use in classifying drowsy conditions when driving in VR simulation, with the highest test accuracy obtained being 91% with the highest precision, sensitivity and F1-score average values of 0.93, 0.91, 0.92 across all classes. The combination of the number of hidden layers and LSTM units that is most optimal for the accuracy of the model is 6 hidden layers with a total of 50 LSTM units. While limited to a small, male-only cohort to ensure optimal EEG electrode contact and minimize noise, this proof-of-concept study successfully demonstrates the feasibility of LSTM in VR environments. Future studies must include a larger, gender-balanced, and more age-diverse cohort of drivers to further validate these findings.

Source of funding: *This study was funded by The Directorate of Research, Technology, and Community Service (DRTPM) on behalf of the Ministry of Education, Culture, Research, and Technology under the BIMA program (Contract: 040/E5/PG.02.00.PL/2024;1754/B/UN3.LPPM/PT. 01.03/2024) and the Faculty Of Science And Technology, Universitas Airlangga. The authors thank them for their financial and technical support throughout this research.*

Author contributions: *Riries Rulaningtyas research concept and design. Sayyidul Istighfar Ittaqillah and Rafi Ihsana Azka Collection and/or assembly of data. Osmalina Nur Rahma and Khouliya Zalda Data analysis and interpretation. Riries Rulaningtyas and Khouliya Zalda Writing the article. Khusnul Ain and Annie Anak Joseph Critical revision of the article. Riries Rulaningtyas and Annie Anak Joseph Final approval of the article.*

Declaration of competing interest: *The author declares no conflict of interest.*

REFERENCE

1. Indonesia BPS. Jumlah Kecelakaan, Korban Mati, Luka Berat, Luka Ringan, dan Kerugian Materi - Tabel Statistik [Internet]. www.bps.go.id. 2024. Indonesia Available (29.06.2026) from: https://www.bps.go.id/id/statistics-table/2/NTEzIzI=/jumlah-kecelakaan-korban-mati-luka-berat-luka-ringan-dan-kerugian-materi.html
2. Rani Rahmadiyani, Ari Widyanti. Prevalence of drowsy driving and modeling its intention: An Indonesian case study. Transportation research interdisciplinary perspectives. 2023; 1;19: 100824–4.
3. Soares S, Monteiro T, Lobo A, Couto A, Cunha L, Ferreira S. Analyzing driver drowsiness: from causes to effects. Sustainability. 2020; 5;12(5):1971.
4. Chen F, Terken J. Automotive Interaction Design. Springer tracts in mechanical engineering. 2023.
5. Gibbings A, Ray LB, Berberian N, Nguyen T, Shahidi Zandi A, Owen AM, Comeau FJE, Fogel SM. EEG and behavioural correlates of mild sleep deprivation and vigilance. Clin Neurophysiol.2021;132(1):45-55. https://doi.org/10.1016/j.clinph.2020.10.010
6. Chai M, Li S-w, Sun W-c, Guo M-z. Drowsiness monitoring based on steering wheel status. Transp Res D Transp Environ. 2018;66:95-104. https://doi.org/10.1016/j.trd.2018.07.007
7. Barua B, Biswas S, Deb K. An efficient method of lane detection and tracking for highway safety. In: 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT). Dhaka, Bangladesh: IEEE; 2019.:1-6. https://doi.org/10.1109/ICASERT.2019.8934664
8. Maheswari VU, Aluvalu R, Kantipudi M P, Chennam K K, Kotecha K, Saini JR. Driver drowsiness prediction based on multiple aspects using image processing techniques. IEEE Access. 2022;10:54980–54990. https://doi.org/10.1109/ACCESS.2022.3176451
9. Wang X, You L, Chen J, Han S. The impact of different service states of tunnel lighting on traffic safety. Accid Anal Prev. 2023;192:107237. https://doi.org/10.1016/j.aap.2023.107237
10. Wang L, Wang H, Liu J. Discrimination of driver fatigue based on distortion energy density theory and multiple physiological signals. IEEE Access. 2021;9:145123-145134. https://doi.org/10.1109/ACCESS.2021.3125052
11. Iwamoto H, Hori K, Fujiwara K, Kano M. Real-driving-implementable drowsy driving detection method using heart rate variability based on long short-term memory and autoencoder. IFAC-PapersOnLine. 2021;54(15):526-531. https://doi.org/10.1016/j.ifacol.2021.10.310
12. Rahma ON, Rahmatillah A. Drowsiness analysis using common spatial pattern and extreme learning machine based on electroencephalogram signal. J Med Signals Sens. 2019;9(2):130-136. https://doi.org/10.4103/JMSS.JMSS_54_18
13. Stephan AM, Lecci S, Cataldi J, Siclari F, et al. Conscious experiences and high-density EEG patterns

predicting subjective sleep depth. Curr Biol. 2021;31(24):5171-5182.e4. https://doi.org/10.1016/j.cub.2021.10.012
14. Filio D, Ziraldo E, Dony L, Gonzalez D, Oliver M. Comparison between wrap around screens and a head-mounted display on driver muscle and kinematic responses to a pedestrian hazard. Appl Ergon. 2023;106:103878. https://doi.org/10.1016/j.apergo.2022.103878
15. Bilgin P, Agres K, Robinson N, Wai AAP, Guan C. A comparative study of mental states in 2D and 3D virtual environments using EEG. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). Bari, Italy: IEEE; 2019: 2833-2838. https://doi.org/10.1109/SMC.2019.8914326
16. Hartfiel B, Stark R. Validity of primary driving tasks in head-mounted display-based driving simulators. Virtual Real. 2021;25:819-833. https://doi.org/10.1007/s10055-020-00496-w
17. Beck M, Pöppel K, Spanring M, Auer A, Prudnikova O, Kopp M, Klambauer G, Brandstetter J, Hochreiter S. xLSTM: Extended long short-term memory. arXiv. 2024. https://doi.org/10.48550/arXiv.2405.04517
18. Kushida CA, Littner M, Morgenthaler T, Alessi C, Bailey D, Coleman Jr J, Hirshkowitz M, Kapen S, Kramer M, Lee-Chiong T, et al. Atlas of clinical sleep medicine. 3rd ed. Philadelphia (PA): Elsevier; 2023. https://doi.org/10.1016/C2018-0-00111-5
19. Gao Z, et al. EEG-based spatio–temporal convolutional neural network for driver fatigue evaluation. IEEE Trans Neural Netw Learn Syst. 2019;30(9):2755-2763 https://doi.org/10.1109/TNNLS.2018.2886414
20. Aidman E, Balin M, Johnson K, et al. Caffeine may disrupt the impact of real-time drowsiness on cognitive performance: a double-blind, placebo-controlled small-sample study. Sci Rep. 2021;11:4027. https://doi.org/10.1038/s41598-021-83504-6
21. Simanjuntak RA, Oesman TI, et al. Analisis hubungan faktor-faktor penyebab kelelahan kerja pada aktivitas pengolahan benih padi. In: Peningkatan Daya Saing Melalui Perbaikan Mutu Produk dalam Rangka Pembangunan Pertanian di Era Industri 4.0. Yogyakarta: Fakultas Agroindustri, Universitas Mercu Buana Yogyakarta; 2020:709–722. Indonesia Available (29.06.2026) from: https://eprints.akprind.ac.id/2402/
22. Phadikar S, Sinha N, Ghosh R. Automatic eyeblink artifact removal from EEG signal using wavelet transform with heuristically optimized threshold. IEEE J Biomed Health Inform. 2021;25(2):475-484. https://doi.org/10.1109/JBHI.2020.2995235

Khouliya ZALDA received the bachelor degree in biomedical engineering from Universitas Airlangga, Indonesia in 2023. In 2024, He also finished the master degree in Biomedical engineering in the same university. His research interests include Biomedical instrumentation and signal processing.
e-mail: khouliya.zalda-2023@fst.unair.ac.id

Riries RULANINGTYAS received her bachelor and master degree from Institut Teknologi Sepuluh Nopember, Surabaya. She received her Ph.D. degree from the School of Electrical Engineering and Informatics majoring in Biomedical Engineering, Institut Teknologi Bandung. Her research interests include medical signal processing, medical image processing, and artificial intelligence. She is currently an associate professor in the Biomedical Engineering Study Program, Department of Physics, Faculty of Science and Technology, Universitas Airlangga, Indonesia.
e-mail: riries-r@fst.unair.ac.id

Khusnul AIN received his bachelor's, master's, and doctoral degrees from Universitas Gadjah Mada, Indonesia majoring in Nuclear Engineering and physics, and Bandung Institute of Technology majoring in Physics Engineering, Indonesia. His research fields are tomography and spectroscopy of electrical bio-impedance. He is currently a senior lecturer in Bachelor of Biomedical Engineering, Department of Physics, Faculty of Science and Technology, Airlangga University, Indonesia.
e-mail: k_ain@fst.unair.ac.id

Osmalina Nur RAHMA received her bachelor degree from Universitas Airlangga, Indonesia, and her master's degree from Universitas Indonesia, Indonesia. Her research field is biomedical product development. She is currently a lecturer in Bachelor of Biomedical Engineering, Department of Physics, Faculty of Science and Technology, Universitas Airlangga, Indonesia.
e-mail: osmalina.n.rahma@fst.unair.ac.id

Sayyidul Istighfar ITTAQILLAH received his bachelor. degree in biomedical engineering from Universitas Airlangga, Indonesia in 2024 and master degree in biomedical engineering from Universitas Airlangga in 2025 specializing in medical instrumentation. He is involved in research on biomedical subject areas with interests in medical signal and image processing, artificial intelligence, and biomaterial engineering.
e-mail: osmalina.n.rahma@fst.unair.ac.id

Annie Anak JOSEPH received her bachelor degree in Electronic and Electrical Engineering from College University Technology Tun Hussein Onn, Malaysia, in 2005. She received her M.S. degree in 2006 from University Science Malaysia. Then, she joins as a lecturer in Universiti Malaysia Sarawak under Department of Electronic, Faculty of Engineering in 2006. She then received Doctor of Engineering in Electrical and Electronic Engineering at Kobe University, Japan and promoted as a senior lecturer in 2014. Her research interests include online learning, concept drift, feature extraction and machine learning.
Email: jannie@unimas.my

Rafi Ihsana AZKA received his bachelor degree in Biomedical Engineering from universitas Airlangga. Currently became a master student in biomedical engineering from Universitas Airlangga specializing in medical instrumentation. He is involved in research on biomedical subject areas with interests in medical device, image processing and aritifical intelligence.
e-mail: rafi.i.a2074@gmail.com

Nabila Nayara FAWZA received her bachelor's degree in Biomedical Engineering from Universitas Airlangga. She is currently pursuing her studies in the field of biomedical engineering, with a strong interest in medical instrumentation. She is actively involved in research in biomedical areas, particularly focusing on medical devices, image processing, and artificial intelligence
e-mail: nabilanayara7@gmail.com