

CLASSIFICATION OF MENTAL FATIGUE BASED ON EEG SIGNAL ANALYSIS USING THE K-NN ALGORITHM

Vinsensia Munte, Tiorina, Abdi Dharma and Mardi Turnip

¹⁾ Information Systems Study Program, Faculty of Science and Technology, Universitas Prima Indonesia, Medan, Indonesia

(Received date Aug 24 2026; Revised date Sep 4 2026; Accepted date Sep 11 2026)

ABSTRACT Mental fatigue is a condition that can impact an individual's productivity and health. Commonly used diagnostic methods are subjective, making it difficult to accurately and in real-time detect a person's level of mental fatigue. Using electroencephalogram (EEG) signals, a technology for measuring brain activity, objective data related to mental fatigue can be obtained. The K-Nearest Neighbor (K-NN) algorithm is used to classify fatigue levels by analyzing patterns formed from EEG signals. EEG was recorded from four participants with a Muse headband (channels TP9, AF7, AF8, TP10) during relaxed, concentrated and neutral sessions while they viewed emotion-eliciting film clips. Class labels ("Inactive", "Less Active", "Active") were assigned from the relative dominance of EEG frequency bands rather than from a validated fatigue instrument, so the task is more precisely described as classification of EEG band-power states than as diagnosis of clinically defined mental fatigue. A tuned K-NN model reached 92% accuracy on a held-out test set and 90.6% under 10-fold cross-validation for the three band-power classes; because validation was performed at the sample level within only four participants, these figures indicate preliminary feasibility rather than a validated, real-time mental-fatigue detector.

KEY WORDS : EEG, K-NN, Kaggle

1. INTRODUCTION

Mental fatigue is a very common condition in modern life, especially experienced by workers, students, and individuals with high cognitive loads (Smith and Doe, 2020). The impact of this condition is extensive, including a decreased ability to concentrate and maintain focus, as well as a significant decrease in productivity. According to a recent report, global economic losses due to stress and mental fatigue are estimated to reach more than 1 trillion US dollars annually. In addition to reducing productivity, mental fatigue also affects safety, especially in the transportation sector. Data from the NHTSA shows that approximately 91,000 traffic accidents are caused by fatigued drivers, with more than 50,000 injuries and 800 deaths (National Highway Traffic Safety Administration, 2026). On the other hand, the diagnosis of mental fatigue still relies on subjective methods, which often lead to under- or over-diagnosis. The absence of definitive physiological indicators causes

many cases to go undetected early or even untreated (Monteiro et al., 2019). The WHO reports that more than 60% of office workers experience symptoms of mental fatigue, but only 20% receive professional treatment (World Health Organization, 2020).

For clarity, this study treats mental fatigue as a gradual decline in the efficiency of sustained attention and cognitive control that follows prolonged mental effort. It is related to but distinct from acute emotion (a short-lived affective response), cognitive workload (the momentary demand a task places on working memory), psychological stress (a response to perceived threat or pressure), drowsiness (reduced arousal and sleep pressure) and transient loss of concentration. These states share EEG correlates - for example increased frontal theta and changes in the alpha band - which makes them easy to confound. This distinction matters here because the film-clip protocol used in this study primarily manipulates emotional and attentional state, and the class labels are derived from EEG band-power dominance rather than from a direct measure of fatigue; the results should therefore be read as discrimination of EEG band-power states that are plausibly associated with fatigue.

Given this reality, a more objective and accurate approach is needed to detect mental fatigue (Park et al., 2023). One way is by using biometric data such as EEG signals. EEG has high potential because it can record brain activity directly, but the signals are complex, non-stationary, and susceptible to noise. EEG signal analysis requires careful and effective data processing and classification techniques to produce accurate information. This study proposes the use of the K-NN algorithm as a classification method because of its simple and efficient nature. K-NN does not require a complex model training process but is still able to provide competitive classification results. This algorithm is very suitable for processing EEG data with limited data amounts but has complex patterns (Kumar and Singh, 2022). This approach is expected to enable a more objective, fast, and accurate diagnosis of mental fatigue. This research is the first step towards the development of a technology-based mental fatigue detection system that is more adaptive and precise.

Various methods have been used to address the problem of EEG signal-based mental fatigue classification (Alotaiby et al., 2015). Support Vector Machine (SVM) is widely used due to its ability to handle non-linear data with kernel techniques (Vapnik, 1998). Artificial Neural Networks (ANN) are also popular due to their ability to recognize complex patterns in EEG signals quite well (Belytschko and Leviathan, 1994). The choice of algorithms such as Decision Tree and Random Forest is based on ease of interpretation and competitive performance (Nainggolan et al., 2024). In terms of feature extraction, Fast Fourier Transform (FFT) is used to analyze EEG signals in the frequency domain (Zhang et al., 2020). These methods often require large datasets and high computational resources to achieve optimal performance. The risk of overfitting is also a serious challenge, especially if the model is not configured properly. These limitations encourage the use of K-NN because this algorithm is lightweight, efficient, and does not require a model training process (Subasi, 2012).

Currently, the latest technologies in EEG signal classification include the use of deep learning architectures such as transformers and attention mechanisms (Shoaib and Alvi, 2021). One breakthrough model is EEG-Former, which applies self-attention to understand the spatial and temporal dynamics of EEG signals (Vaswani et al., 2017). This model has been proven to improve classification accuracy and detect complex mental fatigue patterns (Nugroho et al., 2015). Neuromorphic computing is starting to be developed to process EEG data by mimicking the workings of the human nervous system (Xie et al., 2021). Wearable EEG technology based on edge computing is also being tested to process signals directly on the device without a server (Indiveri and Liu, 2015). These technologies are still in the experimental stage and are not yet commercially available. Limited clinical validation, high computational

requirements, and development complexity are the main obstacles (Kim et al., 2020). Nevertheless, the potential for their application is very large in future real-time mental monitoring systems.

The K-NN algorithm was chosen as a classification method due to its effective ability to recognize patterns in EEG signals. K-NN is a non-parametric algorithm that classifies data based on proximity, generally using Euclidean distance. The classification process is carried out by determining the majority label from a number of K nearest neighbors in the training data. The main advantage of this algorithm lies in its ease of implementation and does not require a complex model training process. Therefore, it is very suitable for the development of initial prototypes or systems with limited computing resources. In its application to EEG signals, K-NN has been proven to be able to classify various psychological conditions such as stress, relaxation, and mental fatigue. With optimal feature extraction and data preprocessing processes, this algorithm demonstrates competitive classification performance. Techniques such as signal transformation and dimensionality reduction can be used to improve the accuracy and efficiency of EEG data processing. In addition, K-NN is supported by various libraries and platforms such as Python and MATLAB that provide ready-to-use classification functions (Xie et al., 2020). This makes its implementation faster and more efficient. With these characteristics, K-NN remains relevant and suitable for use in EEG-based mental fatigue detection research, especially in the early stages of system development.

Recent methods such as EEG-Former and neuromorphic computing offer a more complex approach than K-NN in EEG signal analysis (Roy et al., 2019). EEG-Former uses a transformer architecture with a self-attention mechanism to deeply capture the temporal and spatial relationships of EEG signals (Craik et al., 2019). This model outperforms conventional methods in various EEG classification tasks due to its ability to adaptively extract context features. Meanwhile, neuromorphic computing models brain function with energy-efficient neuromorphic networks and spiking neural networks (SNN) (Yang and Gao, 2019). However, this technology is still in development and requires specialized hardware such as Intel's Loihi chip. In contrast, K-NN is a simpler and technically mature algorithm that works by measuring distances between data (Cover and Hart, 1967). K-NN does not require complex model training, making it ideal for applications with computational constraints. While EEG-Former is more accurate on large datasets, K-NN excels in terms of speed, ease of implementation, and efficiency for small to medium datasets (Bashar and Bhuiyan, 2016).

2. METHODOLOGY

The analysis method includes data pre-processing, K-NN algorithm model development, and evaluation to measure the performance of mental state classification. The data pre-processing stage begins with signal filtering and then feature extraction on the EEG dataset. After the feature extraction process is complete, the transformed data is then processed and classified using the K-NN algorithm (Harris et al., 2020). At this stage, a learning machine is built using the Python programming language with the K-NN algorithm as the main classification model (Pedregosa et al., 2011). This algorithm functions to group data based on the proximity of the test data and training data whose classes have been known (McKinney, 2010). The built K-NN model is then evaluated using evaluation metrics such as accuracy, precision, recall, and F1-score. This evaluation aims to assess the model's performance in classifying mental fatigue conditions based on the temporal and spectral characteristics of the EEG signal (Virtanen et al., 2020).

The process of classifying brain activity levels begins with EEG data collection. Participants are asked to watch emotionally arousing film clips, either scenes that evoke feelings of joy or sadness. The goal is to trigger an affective response that can be recorded through brain signals, resulting in variations in EEG wave activity. While participants watched the clip, brain signal data was recorded using a Muse Headband device with four electrodes. This device functions as a recording device for the brain's electrical activity, placed in the frontal and temporal regions of the head. The resulting raw data is then displayed as a graph of signal activity from each channel. The obtained EEG data still contains noise and physiological artifacts, requiring a signal filtering stage. In this process, digital filtering techniques are used to remove unwanted interference such as signals from muscle movement or electrical interference. The filtering results in a cleaner signal that is ready for further processing. After the signal is cleaned, feature extraction is performed on the EEG waves. This process produces numerical features that represent the level of brain activity in each channel and specific wave type. These features are then used as input for the classification process. The final stage of this process is data classification using the KNN algorithm. This algorithm works by finding the nearest neighbors of new data in the feature space and then determining its class based on the majority class of those neighbors.

This research was conducted at a university in Birmingham, England, in collaboration with Aston University and the Federal University of Paraná. EEG signals were recorded using the Muse Headband through dry electrodes at four placement points. The placement points are at TP9, AF7, AF8, and TP10 as shown in Figure 1, marked with yellow dots and green dots (NZ) as calibration references.

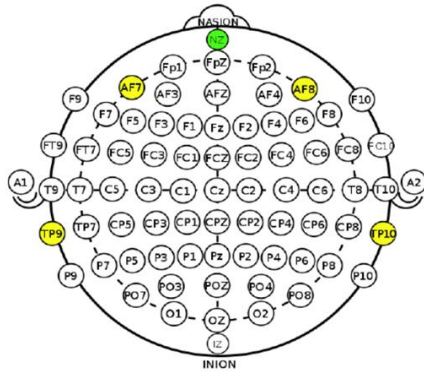


Figure 1. Electrode Placement.

This study involved four participants: two men and two women aged 20-22. Subjects were selected without any specific criteria related to previous experience to obtain representative brain activity data (Salim et al., 2015). Subjects sat in a quiet, comfortable room during the sessions. Recordings were conducted in three mental states: relaxed, concentrated, and neutral. EEG data were recorded and stored using specialized software compatible with Muse EEG, ensuring accuracy and high signal quality.

EEG signals were recorded while subjects watched film clips that triggered emotional responses, including scenes with varying themes of sadness and happiness. As subjects watched the clips, their brain activity changed according to the emotional response they evoked. Mental fatigue is a condition that directly affects concentration, attention, and cognitive abilities. Emotional responses to film clips can accelerate the onset of signs of mental fatigue, especially when individuals are not in optimal rest. Sad movie clips can decrease motivation or induce stress. Meanwhile, overly joyful or complex clips can demand excessive attention and increase cognitive load (Bird et al., 2018). EEG wave patterns such as delta, theta, alpha, beta, and gamma show clear differences between brain states of fatigue and those of non-fatigue watching clips. Each participant underwent several EEG recording sessions in three mental states: relaxed, concentrated, and neutral. Each session lasted 60 seconds, or one minute. In one minute, the EEG data generated thousands of samples from each channel (Nugroho et al., 2023). Each participant underwent multiple sessions, resulting in millions of data points, even though they came from only four subjects. The data was processed to generate

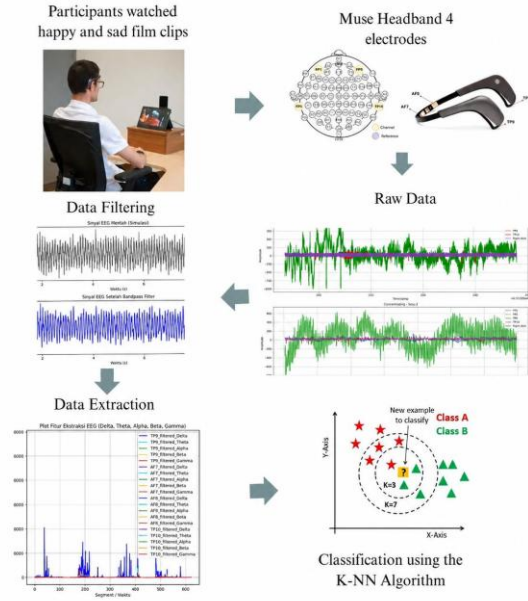


Figure 2. Block Diagram.

3. RESULTS AND DISCUSSION

After 60 seconds of recording, EEG signals were obtained from four electrode channels: TP9, AF7, AF8, and TP10. Each channel represents a different location on the head surface based on the international 10–20 electrode placement system. This is designed to ensure a systematic distribution of electrodes to record electrical brain activity from the frontal, parietal, temporal, and occipital areas. The horizontal (X) axis displays the time in seconds, indicating the duration of data collection. The vertical (Y) axis displays the EEG signals from each channel in amplitude. Each colored line represents the electrical activity from a specific electrode channel. The average amplitude value is listed next to the channel name, as shown in Figure 3. This visualization represents the EEG signals obtained from subjects A, B, C, and D during a recording session in three different mental states: concentration (K), neutral (N), and relaxed (R). Each condition was performed in two separate sessions, allowing for comparison of signal stability and variation over time (Hunter, 2007).

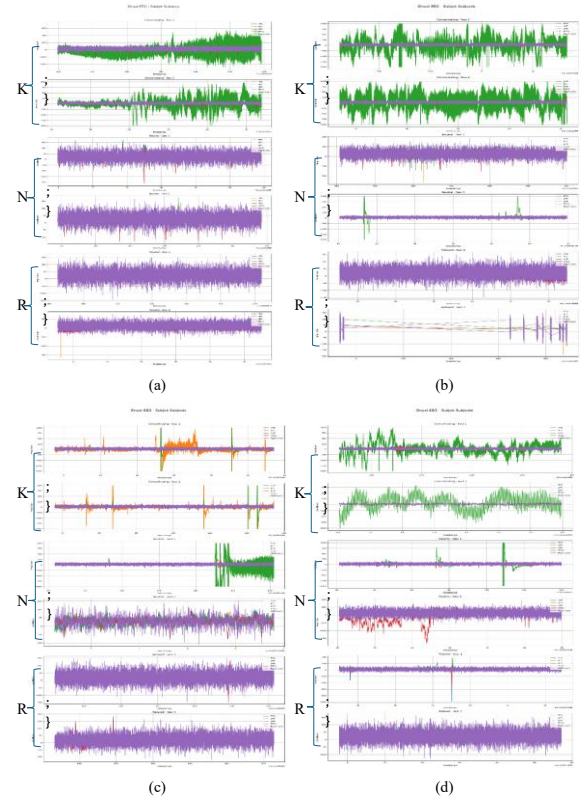


Figure 3. Raw Data (a) Subject A, (b) Subject B, (c) Subject C, (d) Subject D

Figure 3(a) shows a visualization of the EEG signal from subject A. In the K(a) section, the EEG signal shows very large fluctuations, especially in frontal channels such as AF8 and AF7. The amplitude increases significantly, accompanied by sharp spikes indicating intense brain activity. This pattern reflects the condition when the subject is in a state of full focus, where cognitive activity is at a high level. In the N(a) section, which represents a neutral condition, the EEG signal appears more stable with a moderate amplitude and a repetitive wave pattern. Channels such as TP9 and TP10 do not show large spikes, but instead show relatively calm waves, reflecting a brain in a state without dominant emotional or cognitive stimulation. Meanwhile, in the R(a) section, which represents a relaxed state, the EEG signal has a lower amplitude and minimal fluctuations. The recorded waves tend to be slow, leading to the characteristics of alpha and theta waves that usually appear when the body and mind are in a relaxed state or approaching sleep. These differences in signal patterns overall illustrate the dynamics of subject A's brain activity at various levels of consciousness and cognitive engagement.

Figure 3(b) shows the EEG signal recording of subject B. In section K(b), The EEG signal exhibits high amplitude and significant fluctuations, particularly in frontal channels such as AF8 and AF7. This dense, spike-

filled wave pattern indicates a very high level of brain activity, indicating that the subject is in an active and focused cognitive state. This phenomenon is likely related to the dominance of beta or gamma waves, which commonly occur when someone is concentrating intensely on a task. In the N(b) section, representing the neutral condition, the EEG signal shows a decrease in fluctuations with a calmer amplitude compared to the K(b) condition. The frontal and temporal channels show a more stable and relatively flat pattern. However, in the second neutral session, there are disturbances in the form of signal spikes that are inconsistent with the general pattern, indicating the presence of noise or external signal interference. Meanwhile, the R(b) section depicts a relaxed state, where the signal amplitude generally decreases with small variations, indicating that brain activity is in a low state typical of the resting phase. However, in the second relaxed session, the signal shows striking irregularities, both temporally and in amplitude between channels, indicating artifacts or technical interference in the recording process. This overall image reflects the dynamics of subject B's brain activity which varies based on mental condition, with striking differences especially in the concentration condition and the presence of disturbances in the relaxation session.

Figure 3(c) displays the EEG signal from subject C. In the K(c) section, the EEG signal while the subject was in a state of concentration shows quite significant fluctuations, especially in frontal channels such as AF7 and AF8. The first session displays high-amplitude spikes accompanied by occasional sharp dips, reflecting intense mental activity and possible signal interference such as artifacts from head movements or muscle contractions. This pattern continues in the second session, with inconsistent signal extremes appearing in several channels, further strengthening the possibility of noise or technical interference during the recording process. In the N(c) section, which represents the neutral condition, two sessions are visible that exhibit different characteristics. The first session appears quite calm at first, with stable amplitudes, but towards the end of the recording time, sudden spikes appear in channels AF8 and TP10, possibly indicating a state transition or temporary disturbance. In contrast, the second session in the N(c) section shows better stability, with signal variations that are

The presence of wakefulness and the absence of significant spikes reflect the subject's mental state, which is not overly stimulated. In the R(c) section, representing the relaxed state, both sessions show the most stable signal pattern among all conditions. Amplitude fluctuations tend to be small and no major spikes or disturbances are visible. The Right AUX channel appears more visually dominant due to its color intensity, but generally does not indicate any abnormal activity. This pattern reflects a calm brain state and lower neural activity, as is typical when someone is relaxed or at rest.

Overall, the EEG signal of subject C shows a clear difference between the dynamic and disturbed concentration state, compared to the more stable and calm neutral and relaxed state.

Figure 3(d) shows a visualization of the EEG signal from subject D. In the K(d) section, representing the concentration state, a striking amplitude spike is visible, especially in the frontal channel such as AF8. The signal shows extreme fluctuations that last for a short duration but are intense, reflecting highly active brain activity and responsiveness to cognitive tasks. This pattern indicates intense mental work, where the brain experiences an increase in beta or gamma waves in response to high focus. Moving on to the N(d) section, which is the neutral state, the two sessions show slightly different characteristics. The first session is relatively stable, but there is a sudden interruption in the AF8 channel, indicating a transient response or noise. The second session shows a smoother signal pattern, but the TP10 channel records higher fluctuations than the other channels, which may reflect a mild increase in sensory activity or an insignificant emotional reaction (Turnip and Santoso, 2017). Meanwhile, in the R(d) section, depicting the relaxed state, the EEG signal appears more stable and even across almost all channels in the first session, with one sharp interruption from the TP9 channel showing a fairly large negative amplitude spike, possibly an artifact or a momentary response to an external stimulus. The second session in the relaxed state showed higher stability, with the signal dominating from the Right AUX channel, which, despite having a larger amplitude, did not show a significant disturbance pattern. Overall, the EEG signal pattern in subject D showed consistent differences between conditions, where the concentration state was characterized by intense signal dynamics and large amplitudes, while the neutral and relaxed states showed higher signal stability and calmer brain activity patterns.

EEG data obtained from four participants were recorded using a Muse device and saved in.csv format. Each file contains the Timestamp, TP9, AF7, AF8, TP10, and Right AUX columns, representing signals from electrode channels attached to the subject's head. Each file is labeled with metadata such as subject, condition, and session. These wave patterns are then further analyzed to identify correlations between brain activity and the subject's condition. The first step in data processing is pre-processing of the raw EEG signal, which includes filtering to remove noise and unwanted

artifacts. A bandpass filter is applied in the frequency range of 0.5–50 Hz to preserve the main components of brain waves such as Delta (0.5–4 Hz), Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), and Gamma (30–50 Hz). This filtering stage is crucial to ensure that the data to be analyzed is a valid representation of neural activity. Figure 4 visualizes the signal processing process from its raw state to after filtering with a bandpass filter. Plot 4(a)

shows the raw EEG signal from the simulation. This signal exhibits complex amplitude fluctuations and encompasses various frequency components. This raw signal characteristically contains noise irrelevant to the brain activity being observed. The signal appears unstructured, indicating contamination from various sources of interference. Meanwhile, plot 4(b) shows the EEG signal after bandpass filtering. After filtering, the signal appears more structured, smoother, and free from external interference. The amplitude fluctuations become more focused on oscillatory patterns that correspond to brain rhythms, thus better representing actual neural activity.

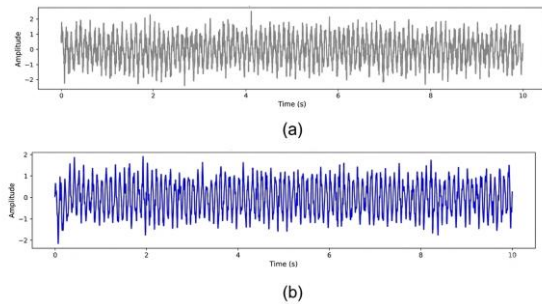


Figure 4. Filtering Data (a) Before, (b) After

After filtering, feature extraction is performed from the EEG signal. This extraction is done by calculating the average spectral power of each frequency band in each EEG channel. This produces features such as the total power of Delta, Theta, Alpha, Beta, and Gamma. The visualization of the EEG signal feature extraction results can be seen in Figure 5. The horizontal axis represents segments or time units in the form of an index, while the vertical axis shows the feature amplitude value or signal strength in each frequency band. From this visualization, it can be observed that the delta wave feature appears to dominate in blue. Where the amplitude value is much higher than the other frequency bands. This indicates that in many segments, delta activity is the dominant component of the EEG signal. This is related to conditions of drowsiness, sleep, or high mental fatigue. Theta, alpha, beta, and gamma waves have lower amplitudes and tend to be stable in most channels, except at certain points that show spikes indicating the possibility of significant cognitive activity or a response to a stimulus.

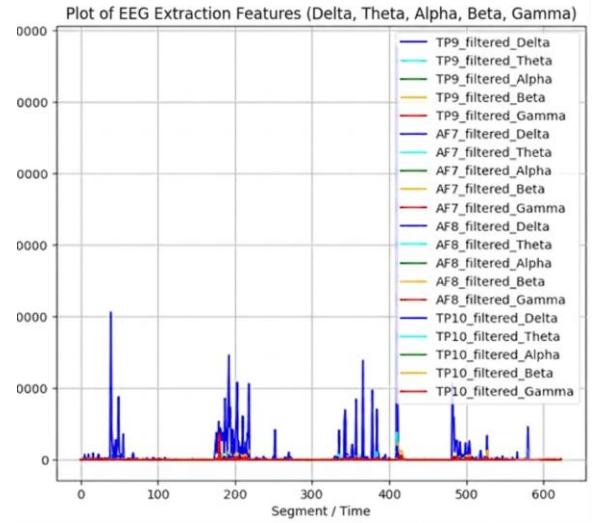


Figure 5. Data Extraction

The extraction results are then summarized into a single data file with a standardized column structure. This includes information on subject identity, data collection conditions, sessions, and feature values for each frequency band. To build a classification system, brain activity is labeled based on the dominance of each frequency feature. If the signal is dominated by Delta and Theta waves, the brain activity is categorized as "Inactive," indicating a high level of mental fatigue. If Alpha and Theta waves dominate, it is categorized as "Less Active," reflecting moderate fatigue or decreased concentration. Meanwhile, if Beta and Gamma waves

dominate, the condition is assumed to be "Active," meaning an optimal cognitive state. After labeling is complete, feature normalization is performed using the Standard Scalerto ensure that each feature has a comparable scale. This normalization is important to avoid the dominance of certain features in calculating distances between points in the KNN algorithm. To improve model efficiency and avoid overfitting, feature selection is carried out using the method Select K Best with the ANOVA F-statistic evaluation function. The 20 best features with the highest correlation to the label were selected for use in the next stage. This selection aims to retaining the most informative features for classification without including redundant or irrelevant features. Initial analysis of the label distribution indicated an imbalance in the amount of data between classes. To address this problem, the technique was used Synthetic Minority Oversampling Technique (SMOTE). Generates synthetic samples from the minority class by interpolating values between nearest neighbors. This step is done to ensure a more even distribution of data within each class, thus ensuring the KNN model is not biased towards the majority class (Panjaitan et al., 2022).

After the data is balanced, the data is separated into two subsets, namely training data and test data, with a proportion of 80:20. Using the method *stratified sampling* to maintain proportional class distribution, the KNN model is trained using varying parameter combinations through a grid search approach to achieve the best classification performance (Rauf et al., 2022). Parameter adjustments include the number of nearest neighbors ($n_neighbors$), the weighting method (uniform or distance), and the type of distance metric such as Euclidean, Manhattan, and Chebyshev. Each of these distance metrics has different calculation characteristics in determining the proximity between data. Distance Euclidean measuring the straight distance between two points in a dimensional space, which is formulated as:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

While the distance Manhattan using the sum of the absolute differences between dimensions:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (2)$$

And distance Chebyshev measure the maximum difference in one dimension:

$$d(x, y) = \max_i (|x_i - y_i|) \quad (3)$$

Function $d(x, y)$ states the distance between two vectors or points x and y . Vector x and y each consisting of a number of components, where n is the number of dimensions or features of the space. By testing various combinations of these parameters through grid search, the model training process can identify the most optimal configuration, thus achieving the highest accuracy in brain activity classification. This strategy is crucial in ensuring that the KNN model can more accurately distinguish between inactive, less active, and active states. The best model obtained from the grid search process.

The search function is then used to make predictions on the test data. The predicted results are compared with the actual labels. The Original Label indicates the actual brain activity category based on predefined physiological rules. The Prediction function shows the model's predictions based on the data. The class distribution based on the original labels shows that the number of data in each category is relatively balanced. That is, 95 samples fall into the Inactive category (label 0), 96 samples are classified as Less Active (label 1), and the remaining 96 samples fall into the Active category (label 2). This balance is important because it provides fair conditions for the model to learn and avoids bias toward any particular class during the training process. The KNN model was trained using normalized data, feature selected, and balanced using the SMOTE technique. The model was then tested on the test set and produced predictions that were generally very close to the original labels. A distribution of predictions close to the proportion of the original labels indicates that the model has stable performance and is not heavily biased toward any particular class. This is especially important in EEG-

based classification systems that require high sensitivity to all classes.

The confusion matrix in Figure 6 shows the performance of the KNN classification model in identifying three categories of brain activity based on EEG signal data. This matrix represents the number of correct and incorrect predictions for each class, thus providing a comprehensive overview of the model's classification ability. In the Inactive class, the model successfully classified 78 samples correctly, while the remaining 6 samples were incorrectly predicted as Less Active and 11 samples as Active. The number of correct predictions indicates that the model has sufficient sensitivity in recognizing activity patterns that correspond to these categories. Although there is still overlap between the Inactive class and the other two classes, the model demonstrated very high classification performance for the Less Active class, where all 96 samples were correctly classified without any misclassifications.

These results demonstrate that the features of the Less Active class have unique characteristics and are easily distinguished by the KNN model, facilitating the identification process. Meanwhile, for the Active class, the model also performed very well, with 91 of the 96 samples correctly classified. Only five samples were mispredicted, with three samples predicted in each as Inactive, and the other two samples were predicted as Less Active. This indicates that the model is capable of recognizing high-activity patterns from EEG signals with a very high degree of accuracy. Although there is still a small possibility of misclassification.



Figure 6. Confusion Matrix

Overall, the classification results depicted in this confusion matrix indicate that the KNN model used has excellent and stable classification performance across all three classes. High classification accuracy and minimal prediction error indicate that the model is capable of recognizing data structures well. Thus, the KNN-based classification approach in this study proved effective in identifying brain activity levels based on available EEG signal features.

After obtaining classification results using the K-NN algorithm on the EEG dataset, the next step was to evaluate the model's performance. Model performance evaluation was performed by calculating key classification metrics such as precision, recall, and f1-score for each class. Based on the classification results, the tuned KNN model achieved an overall accuracy of 92% on the test data. Furthermore, a 10-fold cross-validation evaluation was performed (10-fold cross-validation), which indicates that the model has an average accuracy of 90.63% on the training data. This value indicates that the model not only performs well on a specific subset of the data, but also exhibits stable and consistent performance.

Overall. In more detail, for the Inactive class (label 0), the model achieved a precision of 0.96, meaning 96% of the predictions for this class were correct. However, the recall was only 0.82, indicating that there were still a number of samples from the Inactive class that were not correctly classified. This is also evident from the f1-score value of 0.89, which is the harmonic mean between precision and recall. For the Less Active class (label 1), the model performed very well with a precision of 0.92 and a perfect recall of 1.00. This means that all data that should have been classified as Less Active was correctly recognized by the model, without misclassification into other classes.

The f1-score of 0.96 indicates a strong balance between the model's ability to recognize this class and its prediction accuracy. Performance on the Active class (label 2) is also excellent, with a precision of 0.89 and a recall of 0.95. This indicates that the model is quite reliable in recognizing data with high activity levels. However, a small portion of data from this class is still predicted as another class. The f1-score obtained for this class is 0.92. The macro average and weighted average values of precision, recall, and f1-score all range from 0.92 to 0.93. This reinforces the conclusion that the model has even and balanced performance across classes, with no significant bias toward any particular class. The macro average calculates the average metrics for each class equally, while the weighted average considers the proportion of data per class. This evaluation indicates that the KNN model successfully performed the classification task. The high accuracy values, both in direct testing and cross-validation, indicate that the model has good generalization capabilities to new data. However, the slight decrease in recall for the Inactive class indicates that this class is slightly more difficult to recognize than the other two classes. This could be due to feature similarities between the classes or overlapping data distributions in the feature space.

Overall, this classification approach demonstrates that with appropriate data processing through feature selection, class balancing, and parameter optimization, the K-NN algorithm can provide good classification performance in identifying brain activity levels based on

EEG data. The resulting model has a high and consistent level of accuracy and is reliable for biomedical signal-based classification applications.

4. LIMITATIONS

This study has several limitations. First, only four participants were recorded, which is insufficient to support conclusions about generalisability. Second, the class labels are derived from the relative dominance of EEG frequency bands rather than from a validated fatigue instrument, and the film-clip protocol manipulates emotion and attention as much as fatigue, so the model may be discriminating experimental conditions or affective states rather than mental fatigue as such. Third, resampling and feature selection were applied before a sample-level train/test split within the same participants, which can cause data leakage and optimistic accuracy estimates; subject-independent validation was not performed. Fourth, no baseline classifiers were evaluated, so the benefit of K-NN over simpler or standard alternatives is not established. Fifth, data were collected under a single acquisition setup and protocol. Accordingly, the results should be interpreted as a preliminary feasibility study rather than as a validated real-time mental-fatigue detection system.

5. CONCLUSION

The results show that a K-NN classifier can separate three EEG band-power states (Inactive, Less Active, Active) with 92% accuracy on a held-out test set and 90.6% under 10-fold cross-validation. Because these estimates come from sample-level validation on four participants with labels defined by band-power rules, they demonstrate the preliminary feasibility of K-NN-based EEG classification rather than a validated real-time mental-fatigue diagnostic system. The model successfully recognized three categories of brain activity: Inactive, Less Active, and Active. Where these three activities represent the cognitive state of each individual. These results prove that the EEG-based approach and the K-NN algorithm are effective for classifying EEG signals according to the level of brain activity. Future work should use a validated fatigue reference measure to define ground-truth labels, adopt subject-independent validation on a larger and more diverse sample, compare K-NN against SVM, Random Forest and Logistic Regression baselines, and analyse the EEG frequency-band features most associated with fatigue.

ACKNOWLEDGEMENT

The authors would like to express their deep appreciation to the Faculty of Science and Technology, Prima Indonesia University, and the supervisors for their support and guidance throughout the research process.

REFERENCES

- [1] Smith, J. dan Doe, A. (2020). Mental fatigue in modern society. *Journal of Cognitive Health* 12, 3, 45–60.
- [2] National Highway Traffic Safety Administration (2026). Drowsy driving statistics. <https://www.nhtsa.gov/risky-driving/drowsy-driving>.
- [3] Monteiro, T. G., Skourup, C. dan Zhang, H. (2019). Using EEG for mental fatigue assessment: A comprehensive look into the current state of the art. *IEEE Trans. Hum.-Mach. Syst.* 49, 6, 599–610.
- [4] World Health Organization (2020). Global workplace mental health report.
- [5] Kumar, R. dan Singh, P. (2022). Classification of EEG signals using K-NN algorithm. *Int. J. Biomed. Eng.* 18, 2, 78–85.
- [6] Alotaiby, T. et al. (2015). EEG signal processing for mental fatigue detection: A review. *Biomed. Signal Process.* 18, 1, 1–15.
- [7] Vapnik, V. (1998). *Statistical Learning Theory*. Wiley-Interscience. New York.
- [8] Zhang, Y. et al. (2020). Mental fatigue classification using EEG and machine learning. *IEEE Access* 8, 124304–124312.
- [9] Subasi, A. (2012). EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Syst. Appl.* 38, 7, 8945–8952.
- [10] Shoaib, M. dan Alvi, A. (2021). A comparative study of machine learning algorithms for EEG signal classification. *J. Intell. Syst.* 30, 4, 472–481.
- [11] Vaswani, A. et al. (2017). Attention is all you need. *Adv. Neural Inf. Process. Syst.*
- [12] Xie, Z. et al. (2021). EEG-Former: A transformer-based model for EEG signal classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 1345–1355.
- [13] Indiveri, G. dan Liu, S. C. (2015). Neuromorphic computing: From materials to systems architecture. *Science* 349, 6249, 144–148.
- [14] Kim, J. et al. (2020). Wearable EEG and edge computing: A new paradigm for brain monitoring. *Sensors* 20, 16, 4563.
- [15] Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H. dan Faubert, J. (2019). Deep learning-based electroencephalography analysis: A systematic review. *J. Neural Eng.* 16, 5, 1–119.
- [16] Craik, A., He, Y. dan Contreras-Vidal, J. (2019). Deep learning for electroencephalogram (EEG) classification tasks: A review. *J. Neural Eng.* 16, 3, 031001.
- [17] Vaswani, A. et al. (2017). Attention is all you need. *Adv. Neural Inf. Process. Syst.*
- [18] Yang, H. dan Gao, S. (2019). Neuromorphic computing for brain-inspired artificial intelligence. *Nat. Electron.* 2, 1, 25–35.
- [19] Cover, T. dan Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* 13, 1, 21–27.
- [20] Bashar, S. K. dan Bhuiyan, M. (2016). Classification of motor imagery EEG using deep learning and wavelet features. *Biomed. Signal Process. Control* 29, 37–47.
- [21] Bird, J. J., Manso, L. J., Ribiero, E. P., Ekart, A. dan Faria, D. R. (2018). A study on mental state classification using EEG-based brain-machine interface. *Proc. 9th Int. Conf. Intell. Syst.*
- [22] Rauf, A., Yosrita, E. dan Aziza, R. N. (2022). Klasifikasi sinyal EEG menggunakan model K Nearest Neighbor untuk pengenalan kata yang dibayangkan. *PETIR: J. Pengkajian dan Penerapan Teknik Informatika* 15, 1, 10–19.
- [23] Nainggolan, J. J. et al. (2024). Early detection of drug users from EEG signals using decision tree. *Proc. 2024 IEEE Int. Conf. Control & Automation, Electronics, Robotics, Internet of Things and Artificial Intelligence (CERIA)*, 1–6.
- [24] Pedregosa, F. et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- [25] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 3, 90–95.
- [26] Harris, C. R. et al. (2020). Array programming with NumPy. *Nature* 585, 7825, 357–362.
- [27] McKinney, W. (2010). Data structures for statistical computing in Python. *Proc. 9th Python Sci. Conf.*, 51–56.
- [28] Virtanen, P. et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 3, 261–272.
- [29] Waskom, M. L. (2021). Seaborn: Statistical data visualization. *J. Open Source Softw.* 6, 60, 3021.
- [30] Xie, C., Wang, J. dan Li, Y. (2020). EEG signal classification using KNN and normalization techniques for brain-computer interfaces. *Int. J. Biomed. Eng. Technol.* 34, 4, 297–310.
- [31] Salim, R. I. R., Turnip, A. dan Joelianto, E. (2015). EEG-based brain-controlled wheelchair with four different stimuli frequencies. *Proc. Int. Seminar on Intelligent Technology and Its Applications (ISITIA)*, 261–266.
- [32] Nugroho, H. A., Turnip, A. dan Joelianto, E. (2015). Classification of mental task EEG records using k-nearest neighbor algorithm. *Proc. Int. Conference on Electrical Engineering and Informatics (ICEEI)*, 500–504.
- [33] Nugroho, S. A., Rahmadi, R. dan Turnip, A. (2023). The comparison between EEG channel selection using filter method and wrapper method. *Proc. Int. Conf. on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, 44–48.
- [34] Turnip, A. dan Santoso, M. D. (2017). EEG-based emotion recognition using convolutional neural

network. Internetworking Indonesia Journal 9, 1, 15–19.

- [35] Panjaitan, W. A., Turnip, M. dan Turnip, A. (2022). Classification of imagined speech EEG using k-nearest neighbor. IOP Conf. Series: Materials Science and Engineering.