Sleepiness Detection Using Machine Learning Models on EEG Data

Driver sleepiness is a major cause of road accidents, necessitating effective detection systems to improve safety. This study investigates the use of machine learning (ML) models to automate the detection of driver sleepiness through electroencephalography (EEG) data collected in simulated environments. Various ML models, such as Random Forests (RF), Decision Trees (DT), Logistic Model Trees (LMT) and two ensemble methods (bagging and stacking), were evaluated using 10-fold cross-validation. More specifically, the selected classifiers were trained and tested using EEG data acquired via the MindSet device, including band power, attention, and mediation features to effectively differentiate between "sleepy" and "non-sleepy" subjects. The bagging approach demonstrated superior performance among the classifiers, achieving 74.9% accuracy, 0.749 precision, 0.750 recall, and an Area Under the ROC Curve (AUC) of 0.835.

CCS Concepts: • **Human Computer Interaction → Brain-Computer Interface**; • **Signal Processing → EEG**; • **Artificial Intelligence → Machine Learning**.

Additional Key Words and Phrases: Sleepiness Detection, EEG, Machine Learning

ACM Reference Format:

. 2024. Sleepiness Detection Using Machine Learning Models on EEG Data. In *The 13th EETN Conference on Artificial Intelligence (SETN 2024), September 11–13, 2024, Piraeus, Greece.* ACM, New York, NY, USA, 4 pages.

1 INTRODUCTION

Sleepiness (or interchangeable drowsiness) is a significant concern for drivers, as it can impair reaction times, decision-making, and overall cognitive function, leading to an increased risk of accidents. Various factors contribute to sleepiness while driving, including sleep deprivation, time of day, monotony of the driving task, and sleep disorders. To address these issues, researchers have been exploring different methods for detecting Sleepiness in drivers. One method centres on the driver's behaviour, specifically analyzing actions such as yawning, eye closing and blinking, and head pose, among other similar movements [2]. Another method that has shown promise for assessing sleepiness involves systems that utilize physiological characteristics, including EEG. By monitoring brainwave patterns, EEG can detect changes in brain activity associated with sleepiness. Several key findings in recent studies [9] have associated sleepiness in contexts like driver monitoring with specific EEG bands due to their ability to capture changes in brain activity. Increased theta (4-8 Hz) activity and decreased alpha (8-13 Hz) activity are often indicators of sleepiness. Using EEG-based systems integrated into vehicles, the driver's brain activity could continuously be monitored in real-time. These systems can improve road safety by detecting signs of sleepiness and issuing warnings to alert the driver to take a break or rest. Additionally, advancements in machine learning algorithms allow for the development of more accurate and reliable sleepiness detection systems based on EEG data. Combining EEG (electroencephalography) with ML techniques offers a promising approach to detecting sleepiness in drivers [4].

In this work, we utilized a publicly available EEG-based dataset, characterized by features including the mean power of five EEG frequency bands at the frontoparietal channel Fp1, and two additional features representing attention and

Table 1. Statistical description of the Dataset.

Features	Mean±std	Min	Max
Attention	48.22±21.88	1	100
Meditation	56.58±19.07	1	100
delta	518771.39±599783.66	216	3598743
theta	136242.00±217550.08	138	3194358
low-alpha	33413.34±51970.93	32	699008
high-alpha	30580.76±52681.86	9	785947
low-beta	25640.07±37159.04	2	595549
high-beta	23276.98±43921.12	3	443589
low-gamma	8117.13±15522.90	6	289281
high-gamma	208725.81±329707.32	47	2328370

meditation. The task of sleepiness detection was framed as a binary classification problem, which we approached using tree-based classifiers and two ensemble approaches, stacking and bagging. Our goal was to identify a model with high sensitivity (namely, recall) and effective differentiation between sleepiness states (Sleepy and Non-sleepy) measured by the AUC metric. Using 10-fold cross-validation, the bagging method outperformed the rest models, achieving an accuracy of 74.99%, precision of 0.749, recall of 0.750, and an AUC of 0.835.

The rest of the paper is structured as follows: Section 2 provides a detailed description of the dataset and the methodology employed. In Section 3, we analyze the performance results of the machine learning models. Finally, Section 4 presents the conclusions and potential future research directions.

2 METHODOLOGY

In this section, we present an overview of the dataset and its attributes, outline the methodology employed, describe the machine learning models used, and specify the evaluation metrics applied for the experimental assessment.

2.1 Data Description and Analysis

More specifically, EEG signal data were derived from 4 potential drivers while they were awake and asleep using the NeuroSky MindWave sensor. For safety precautions, they weren't driving while acquiring the signals. Each driver wore the helmet for 5-8 minutes for each label (sleepy, not sleepy) and the signals were acquired approximately every second. The signals are measured in units of microvolts squared per hertz ($\mu V^2/Hz$). This is a measure of the power of the EEG signal at particular frequency bands; δ (0.5-4 Hz), θ (4-8 Hz), α (8-13 Hz)β (13-30 Hz) and γ (30-100 Hz) [10]. The high values are likely since the MindWave sensor is measuring EEG data from a single location on the forehead. The NeuroSky mindwave headset is single-channel and measures the voltage between an electrode resting on the frontal lobe (forehead) and two electrodes (1 ground and 1 reference), each in contact with one earlobe.

Table 1 sums up statistical details of the EEG features in the dataset consisting of 2135 Non-Sleepy and 1600 Sleepy drivers. To assess the attention and mediation levels of drivers, we referred to the MindSet device manual[1]. Specifically, we used NeuroSky's proprietary eSense algorithm to characterize mental states. Initially, NeuroSky ThinkGear technology processes the raw brainwave signals to eliminate ambient noise and muscle movement. Then, the eSense algorithm interprets the eSense meter values, which indicate different ranges of activity. These values are presented on a relative eSense scale from 1 to 100, with categories defined as follows: i) 1-20: Strongly Reduced, ii) 20-40: Reduced, iii) 40-60:

[1]https://developer.neurosky.com/docs/doku.php?id=mindset_instruction_manual

Neutral, iv) 60-80: Slightly Elevated, and v) 80-100: Elevated. Figures 1, according to the previous scaling, show the attention and mediation levels of Sleepy and Non-Sleepy drivers indicating no significant differences. This finding highlights the necessity for incorporating diverse features, not just from EEG data, assuming also from eye tracking [11].

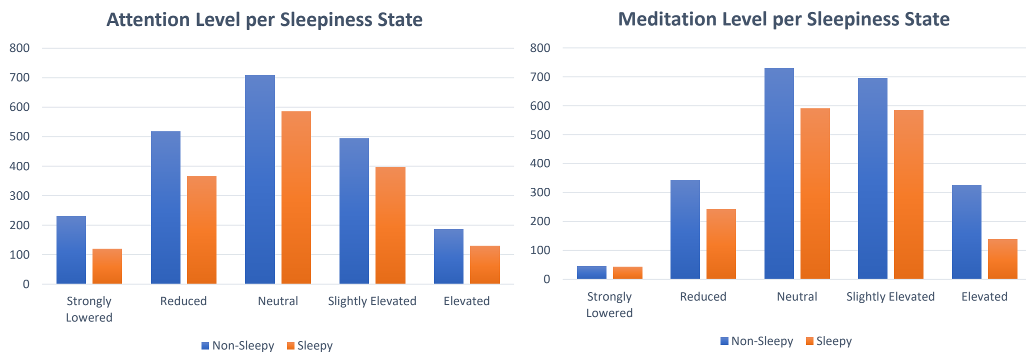

Fig. 1. Attention and Mediation Level as measured by MindSet device.

2.2 ML Models and Evaluation Metrics for Sleepiness Detection

We evaluated our ML models using the free software tool WEKA [1], by assessing some well-known tree-based classification models, and, specifically, RF [3], LMT [5] and J48 DT [7]. Also, for improved performance, we exploited the benefits of bagging which was built upon the J48 and, finally, the stacking ensemble by combining the strengths of the single RF and J48 as base models using logistic regression as the meta-learner [8]. The performance of the classification models was evaluated using appropriate metrics [6], such as accuracy, precision, recall (or sensitivity), and AUC to evaluate the separation ability of a model. Ten-fold cross-validation was employed to assess the models' performance. The overall score in each metric is derived by averaging the results from all folds. Confusion matrix is the basis for defining these metrics consisting of the elements true positive (Tp), true negative (Tn), false positive (Fp) and false-negative (Fn): $\mathrm{Acc} = \frac{\mathrm{Tn+Tp}}{\mathrm{Tn+Fn+Tp+Fp}}$, $\mathrm{Prec} = \frac{\mathrm{Tp}}{\mathrm{Tp+Fp}}$ and $\mathrm{Recall} = \frac{\mathrm{Tp}}{\mathrm{Tp+Fn}}$.

3 RESULTS ANALYSIS

This section presents and analyzes the performance behaviour of the selected ML models and methods, trained and tested in all available features of the dataset. Table 2 shows the mean performance of metrics obtained using 10-fold cross-validation.

Table 2. Average performance outcomes of ML models after 10-fold cross-validation.

	Accuracy	Precision	Recall	AUC
RF	69.45%	0.692	0.696	0.755
LMT	70.52%	0.703	0.705	0.770
J48	72.77%	0.727	0.728	0.766
Stacking (RF, J48 with LR)	73.33%	0.732	0.733	0.806
Bagging (J48)	74.99%	0.749	0.750	0.835

For this dataset, when comparing the models in terms of accuracy, recall, precision and AUC, the RF classifier exhibited the lowest performance. On the other hand, the Bagging method (based on J48) outperformed all investigated models. Also, it is observed that Stacking noted higher improvements in prediction metrics in relation to the RF classifier while its behaviour was more proximal (but still higher) to J48. LMT performance (that combines decision tree learning with logistic regression) was superior to RF with a small performance rise in all metrics. The Bagging classifier was the most efficient at predicting sleepy/non-sleepy drivers, as it maintained the lowest false positive and false negative rates, resulting in fewer misclassifications when compared to rest tree models. The AUC of Bagging indicates a higher probability of 83.5% to correctly distinguish between sleepy and non-sleepy subjects.

4 CONCLUSIONS

To summarize, this study demonstrates the potential of supervised ML for EEG-based sleepiness detection in drivers. The findings suggest that incorporating such models into driver monitoring systems could enhance road safety by providing timely warnings and interventions for drowsy drivers. A limitation of this dataset is the lack of access to the raw data from the Fp1 channel, for extracting additional EEG features in the time or time-frequency domains. Despite this, the models noted satisfactory forecasting capabilities concerning the sleepiness state detection. From the analysis, we concluded that the Bagging method was the most effective, achieving an accuracy of 74.99%, precision of 0.749, recall of 0.750 and an AUC of 0.835. We plan to apply feature selection techniques to comprehend the features' relatedness and significance to the sleepiness detection task and reevaluate the performance of the ML models. Additionally, we aim to develop personalized sleepiness detection models and compare their performance to the current global models. Finally, we will explore different EEG-based datasets and those based on the eye features for further research [11].

REFERENCES

[1] (accessed on 19 May 2024). Weka. https://www.weka.io/.

[2] Afef Abidi, Khaled Ben Khalifa, Ridha Ben Cheikh, Carlos Alberto Valderrama Sakuyama, and Mohamed Hedi Bedoui. 2022. Automatic detection of drowsiness in EEG records based on machine learning approaches. *Neural Processing Letters* 54, 6 (2022), 5225–5249.

[3] Sotiris Alexiou, Elias Dritsas, Otilia Kocsis, Konstantinos Moustakas, and Nikos Fakotakis. 2021. An approach for personalized continuous glucose prediction with regression trees. In *2021 6th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)*. IEEE, 1–6.

[4] Islam A Fouad. 2023. A robust and efficient EEG-based drowsiness detection system using different machine learning algorithms. *Ain Shams engineering journal* 14, 3 (2023), 101895.

[5] Bita Ghasemkhani, Reyat Yilmaz, Derya Birant, and Recep Alp Kut. 2023. Logistic Model Tree Forest for Steel Plates Faults Prediction. *Machines* 11, 7 (2023), 679.

[6] Mohammad Hossin and Md Nasir Sulaiman. 2015. A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process* 5, 2 (2015), 1.

[7] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, and Jonathan Taylor. 2023. Tree-based methods. In *An Introduction to Statistical Learning: with Applications in Python*. Springer, 331–366.

[8] Ammar Mohammed and Rania Kora. 2023. A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University-Computer and Information Sciences* 35, 2 (2023), 757–774.

[9] Igor Stancin, Mario Cifrek, and Alan Jovic. 2021. A review of EEG signal features and their application in driver drowsiness detection systems. *Sensors* 21, 11 (2021), 3786.

[10] Maria Trigka, Elias Dritsas, and Christos Fidas. 2022. A Survey on Signal Processing Methods for EEG-based Brain Computer Interface Systems. In *Proceedings of the 26th Pan-Hellenic Conference on Informatics*. 213–218.

[11] Zainah Md Zain, Mohd Shahril Roseli, and Nurul Athirah Abdullah. 2023. Enhancing Driver Safety: Real-Time Eye Detection for Drowsiness Prevention Driver Assistance Systems. *Engineering Proceedings* 46, 1 (2023), 39.