

12º CONGRESSO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO EM PETRÓLEO E GÁS



TÍTULO DO TRABALHO:

Monitoramento da Sonolência na Indústria de Óleo e Gás: Uma Abordagem com Algoritmos de Aprendizagem de Máquina

AUTORES:

Mateus de Miranda Sotero^{1,2*}, Caio Bezerra Souto Maior^{1,3}, Plínio Marcio da Silva Ramos^{1,2}, Isis Didier Lins^{1,2}, Márcio José das Chagas Moura^{1,2}

INSTITUIÇÃO:

¹Centro de Estudos e Ensaios em Risco e Modelagem Ambiental, ²Departamento de Engenharia de Produção, UFPE, ³Núcleo de Tecnologia, UFPE

MONITORAMENTO DA SONOLÊNCIA NA INDÚSTRIA DE ÓLEO E GÁS: UMA ABORDAGEM COM ALGORITMOS DE APRENDIZADO DE MÁQUINA

Resumo

As salas de controle são essenciais em setores como óleo e gás, operando em refinarias, plataformas *offshore* e *onshore*, e terminais de armazenamento. Elas são centrais para decisões críticas que garantem operações seguras e eficientes. Este trabalho explora como a sonolência compromete a resposta em situações críticas e apresenta ferramentas para mitigar esse risco, abordando a fadiga humana. O estudo propõe uma abordagem utilizando dados de eletroencefalograma (EEG) e algoritmos de aprendizado de máquina para identificar o estado de atenção dos participantes. Usando o banco de dados "ULg multimodality sleepiness database" (DROZY), que contém dados relacionados à sonolência, foram extraídas características específicas dos sinais de EEG, como dimensão fractal, complexidade de Hjorth e mobilidade de Hjorth. A detecção de sonolência foi realizada com modelos clássicos de aprendizado de máquina supervisionado, incluindo Random Forest (RF), Multilayer Perceptron (MLP), K-Nearest Neighbours (KNN) e Support Vector Machine (SVM), além de modelos ensemble como XGBoost e LightGBM. Esses modelos foram treinados e testados usando os dados do DROZY para avaliar sua eficácia na identificação de sonolência. Além disso, o VotingClassifier, um metaclassificador que combina modelos de aprendizado de máquina utilizando um método de votação, foi empregado para equilibrar as fraquezas dos classificadores individuais. Os resultados mostraram que a abordagem é capaz de detectar com precisão a sonolência com base nos dados de EEG, alcançando acurácias acima de 90% para sujeitos específicos. Modelos como XGBoost e LightGBM demonstraram desempenho promissor com acurácias superiores aos outros classificadores. Este estudo oferece uma base sólida para futuras implementações práticas, mostrando que o aprendizado de máquina pode analisar dados de EEG e detectar sonolência de operadores com precisão, contribuindo para a segurança nas operações de energia e mitigando o risco de erros humanos.

Abstract

Control rooms are essential in sectors such as oil and gas, operating in refineries, offshore and onshore platforms, and storage terminals. They are central to critical decisions that ensure safe and efficient operations. This work explores how drowsiness compromises response in critical situations and presents tools to mitigate this risk, addressing human fatigue. The study proposes an approach using electroencephalogram (EEG) data and machine learning algorithms to identify the attention state of participants. Using the "ULg multimodality sleepiness database" (DROZY), which contains data related to drowsiness, specific features were extracted from the EEG signals, such as fractal dimension, Hjorth complexity, and Hjorth mobility. Drowsiness detection was performed using classical supervised machine learning models, including Random Forest (RF), Multilayer Perceptron (MLP), K-Nearest Neighbours (KNN), and Support Vector Machine (SVM), as well as ensemble models like XGBoost and LightGBM. These models were trained and tested using the DROZY data to evaluate their effectiveness in identifying drowsiness. Additionally, the VotingClassifier, a meta-classifier that combines machine learning models using a voting method, was employed to balance the weaknesses of individual classifiers. The results showed that the approach can accurately detect drowsiness based on EEG data, achieving accuracies above 90% for specific subjects. Models such as XGBoost and LightGBM demonstrated promising performance with accuracies surpassing those of the other classifiers. This study provides a solid foundation for future practical implementations, showing that machine learning can analyze EEG data and accurately detect operator drowsiness, contributing to safety in energy operations and mitigating the risk of human error.

Introdução

A perfuração offshore e as operações de energia são processos altamente industrializados e complexos que envolvem vários equipamentos, tarefas complicadas e uma força de trabalho diversificada [1]. Monitorar informações disponíveis e emitir alertas sobre os riscos potenciais de desastres geralmente

são realizados em uma sala de controle [2]. Essas salas são manuseadas por operadores para analisar variáveis ambientais, identificar possíveis perigos e vulnerabilidades, e comunicar alertas com uma equipe de resposta a emergências [3], [4]. No entanto, a ocorrência de erros humanos, como demonstrado pelo incidente da BP Texas City, que uma das causas foi a fadiga de operadores, além de falhas operacionais [5], ressalta a necessidade de sistemas mais avançados de detecção e prevenção de sonolência, que se configura como uma consequência da fadiga humana e uma das causas da desatenção. Além disso, foram estabelecidas não apenas para o gerenciamento de riscos de desastres, mas também para áreas de interesse, como usinas nucleares e refinarias de petróleo[6].

A sonolência representa o estado de transição entre o despertar e o sono, durante o qual ocorre um decréscimo na vigilância. Como antecedente do estado de sono, essa fase é caracterizada por uma quase total falta de reflexos, tornando-se incompatível com tarefas que exigem extrema concentração, rastreamento da sonolência, portanto, é um objetivo importante no aprimoramento da segurança em prol da saúde humana [7]. A sonolência é frequentemente causada pelo desalinhamento do ritmo circadiano interno devido ao trabalho em turnos, privação de sono e atividade física ou mental prolongada [8]. Dessa forma, atividades na indústria de energia necessitam de extrema atenção uma vez que a falta de concentração e a sonolência podem gerar efeitos severos.

O reconhecimento de sonolência baseado na interpretação humana geralmente requer a intervenção de outra pessoa para identificar eventos de fadiga, o que pode estar sujeito a erros; alternativamente, um esquema de sonolência assistida por computador pode ser empregado [9]. Assim, a avaliação do nível de sonolência do operador é feita por meio do uso de visão computacional (por exemplo, imagens/vídeos) ou através de informações biológicas (por exemplo, eletroencefalograma - EEG) [10]. O EEG é um método de monitoramento eletrofisiológico que é utilizado para registrar a atividade elétrica do cérebro, sendo particularmente relevante na identificação de estados críticos como a sonolência[11]. Como ressaltado por Venkata Phanikrishna et al., (2023), a sonolência durante atividades laborais emergiu como uma área de pesquisa essencial nos últimos anos. A prevenção de acidentes causados pela sonolência em sistemas críticos pode ser alcançada por meio da detecção de mudanças nos estados de alerta para sonolência em indivíduos. Neste cenário, os padrões eletroencefalográficos (EEG) oferecem informações biomédicas valiosas para analisar essas alterações nos estados neurológicos [11].

Nesse contexto, o estudo propõe uma análise de desempenho com algoritmos de aprendizado de máquina e aprendizado profundo, utilizando conjuntos de características adquiridos automaticamente a partir de EEG, sendo fundamental para a automatização do diagnóstico desses estados. O restante do artigo é apresentado da seguinte maneira: a Seção 2 trata da Metodologia, na Seção 3 temos os Resultados e Discussões, na Seção 4 a Conclusão e por fim na seção 5 os Agradecimentos.

Metodologia

A compreensão aprofundada da sonolência e suas implicações na atenção e tomada de decisão é crucial em setores de alta demanda cognitiva, como na indústria de petróleo e gás [6], como o DROZY é um benchmarking de estudo para a atenção/sonolência, ele é consistente para ser utilizado na validação de modelos desenvolvidos para detecção destes estados (Ramos et al., 2022). Neste caso, o experimento registrou os canais Fz, Pz, Cz, C3 e C4 amostrados a uma taxa de 512 Hz. No total, quatorze sujeitos/participantes (3 homens, 11 mulheres) com uma idade média de 22,7 anos e um desvio padrão de 2,3 anos foram considerados [13].

Para realizar a análise de detecção de sonolência utilizando machine learning e dados de EEG, foram coletados registros de cinco pessoas (sujeitos 1, 5, 6, 8 e 10) em dois estados distintos: alerta e sonolento do banco de dados DROZY. A escolha desses sujeitos teve como base a sua alternância entre a classificação de alerta para sonolento desde o primeiro PVT até o último [9]. A leitura e manipulação dos dados foram realizadas utilizando a biblioteca MNE no ambiente Python. Para cada sujeito, os dados em estado de alerta foram extraídos de arquivos EFD identificados como (sujeito)-1.edf, enquanto os dados em estado de sonolência foram extraídos de arquivos EFD identificados como (sujeito)-3.edf. Para simplificar a organização dos dados, 2400 intervalos de 1 segundo (cada um com 128 pontos) foi realizado. Após o janelamento dos dados, houve a extração de características relevantes dos registros EEG para alimentar o modelo de machine learning. As características extraídas têm como objetivo refletir a atividade fisiológica ocorrendo dentro do cérebro. A diversidade de padrões e a complexidade da atividade cerebral tornam desafiadora a criação de uma abordagem única e universal

para a extração de características nesse contexto [11]. Essas características extraídas foram escolhidas com base em sua relevância para a detecção de sonolência, e são agora prontas para serem utilizadas como entradas no modelo de machine learning. As características foram extraídas de cada um dos 2400 intervalos de cada canal de cada sujeito em ambos os estados, transformando os 128 dados em uma tupla com 3 informações como indicado na *Figura 1*.

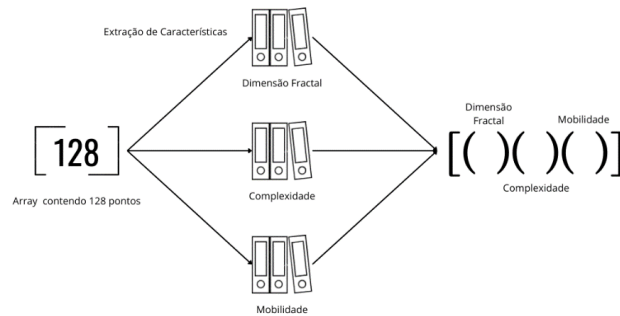


Figura 1: Extração de Características

No contexto deste estudo, a biblioteca EEGLIB [14], foi utilizada para a extração de características dos sinais de eletroencefalograma (EEG). A escolha da EEGLIB fundamentou-se na sua variedade de ferramentas especializadas, proporcionando uma análise precisa e abrangente dos dados EEG.

- **Dimensão Fractal:** Calculada utilizando o algoritmo Higuchi Fractal Dimension [15]. Este parâmetro fornece insights sobre a complexidade e irregularidade dos padrões temporais presentes nos dados.
- **Complexidade de Hjorth:** Este parâmetro avalia as alterações neurofisiológicas em termos de frequência [9], [16].
- **Mobilidade de Hjorth:** Ela está associada à rapidez com que as amplitudes dos sinais mudam ao longo do tempo, proporcionando informações sobre a dinâmica dos registros EEG. É expressa como uma razão por unidade de tempo e pode ser concebida como uma frequência média [9], [17].

Para realizar a detecção de sonolência a partir das características extraídas, diferentes modelos de machine learning foram considerados, cada um com suas próprias características e capacidades. A escolha dos modelos foi baseada na natureza do problema e na capacidade de lidar eficientemente com dados de EEG. As bibliotecas *scikit-learn* [18] e *xgboost* foram utilizadas para implementar e avaliar esses modelos no ambiente Python.

As características extraídas dos dados processados de cada canal de cada sujeito, em estado alerta e sonolento foram agrupadas e divididas em dados de treino (80%) e teste (20%). Os rótulos representam uma escala binária, com 0 para estado sonolento e 1 para estado alerta. A escolha de múltiplos modelos visa explorar diferentes abordagens e avaliar qual deles oferece a melhor acurácia na detecção de sonolência com base nos dados EEG com características extraídas

- **Random Forest:** A técnica Random Forest é uma abordagem de árvore de regressão que emprega a técnica de agregação bootstrap e à randomização de preditores para alcançar uma elevada precisão preditiva [19]. Durante a condução da pesquisa, foram adotados os parâmetros padrões disponíveis na implementação do scikit-learn. O modelo Random Forest foi escolhido devido à sua capacidade de lidar bem com características não lineares e fornecer uma estimativa de importância para cada característica.
- **MLP (Multilayer Perceptron):** MLP é uma das redes mais populares que contém múltiplas camadas sucessivas, sendo a camada de entrada, uma ou mais camadas ocultas e uma camada de saída [20]. A informação é propagada da camada de entrada para a camada de saída por meio das camadas ocultas, e os pesos da rede são atualizados durante a fase de treinamento [21]. A rede neural MLP foi incluída para capturar padrões complexos nos dados, considerando a natureza não linear das relações nos registros EEG.
- **SVM (Support Vector Machine):** Enquanto abordagens clássicas são projetadas para minimizar erros no conjunto de dados de treinamento, o SVM (Support Vector Machine) é baseado no princípio da minimização do risco estrutural enraizado na teoria estatística de aprendizado [22]. O SVM tem como objetivo construir um hiperplano ótimo que maximiza a margem entre as classes e lida bem com espaços de características extensos [9], [23]. A Máquina de Vetores de

Suporte (SVM) foi selecionada por sua eficácia em lidar com dados de alta dimensionalidade e sua capacidade de encontrar hiperplanos de separação complexos.

O aprendizado em conjunto (*Ensemble Learning*), que é um tipo de aprendizado de máquina, é uma combinação de múltiplos modelos de classificação em vez de usar um modelo individual, com o objetivo de resolver um problema específico [24]. Os modelos ensemble também podem ser utilizados na automatização da detecção da sonolência através de sinais de EEG [9].

Nesse artigo foram empregados três modelos ensemble: Extreme Gradient Boosting (XGBoost), Light Gradient Boosting (Light GBM) e VotingClassifier. Cada um desses modelos foi utilizado com o intuito de detectar sonolência. O objetivo é explorar um novo contexto através de métodos acessíveis e adequados para aplicação em ambientes com recursos de hardware limitados, mas que ainda assim proporcionem um desempenho satisfatório.

- **LightGBM:** O algoritmo LightGBM é um tipo de GBDT (Gradient Boosting Decision Tree), utiliza árvores de decisão peça-por-peça e funções de perda aproximadas com aproximação de Taylor de segunda ordem em cada etapa. Em seguida, ele treina uma árvore de decisão para minimizar a segunda ordem de aproximação, o que é análogo ao método de Newton [25]. O LightGBM foi escolhido por sua eficiência e escalabilidade em grandes conjuntos de dados, bem como sua capacidade de lidar com características não lineares.
- **XGBoost:** XGBoost é a abreviação de "eXtreme Gradient Boosting package". É uma implementação eficiente e escalável do framework de gradient boosting proposto por Friedman (2001) [26]. O pacote inclui um solucionador eficiente para modelos lineares e um algoritmo de aprendizado de árvores, ele é projetado para ser extensível, permitindo que os usuários definam facilmente seus próprios objetivos [27]. XGBoost foi escolhido por sua capacidade de lidar com dados desbalanceados e sua eficácia geral em problemas de classificação.
- **VotingClassifier:** A combinação dos modelos individuais foi realizada por meio de uma abordagem de votação, este classificador é um metaclassificador para combinar modelos de aprendizado de máquina iguais ou conceitualmente diferentes, para previsão por meio de votação majoritária [28]. Um modelo ensemble de votação pode equilibrar as fraquezas individuais dos classificadores envolvidos [9].

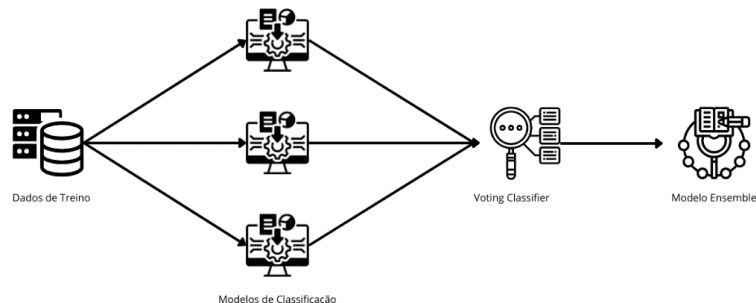


Figura 2: Modelo Ensemble

O conceito por trás de um classificador por votação é combinar diferentes classificadores de aprendizado de máquina e utilizar um critério de votação para prever a classificação [29]. Este ensemble incluiu todos os modelos mencionados anteriormente, e cada classificador contribui com um voto para a classe prevista, como mostrado na Tabela 1 no parâmetro *voting*. A classe final prevista é determinada pela maioria dos votos.

Tabela 1. Arquitetura do modelo ensemble

Algoritmo Ensemble
VotingClassifier (
Random Forest
Multilayer Preceptron
Support Vector Machines
Extreme Gradient Boosting
Light Gradient Boosting
)
voting = hard

Resultados e Discussões

Os avanços na tecnologia de machine learning têm proporcionado oportunidades sem precedentes para a análise e interpretação de dados complexos em uma variedade de campos [30]. Nessa seção, exploramos a acurácia dos modelos de machine learning na classificação de dados obtidos a partir de cinco sujeitos diferentes.

Tabela 2: Acurácia dos Modelos

Sujeitos	Canais	Acurácias					
		Random Forest	MLP	SVM	XGBoost	LightGB	Ensemble
1	Fz	0,53	0,54	0,57	0,55	0,55	0,55
	Cz	0,53	0,53	0,53	0,54	0,53	0,53
	C3	0,61	0,63	0,63	0,62	0,63	0,63
	C4	0,77	0,78	0,79	0,77	0,79	0,8
	Pz	0,54	0,58	0,53	0,53	0,58	0,56
5	Fz	0,6	0,64	0,63	0,6	0,64	0,64
	Cz	0,53	0,54	0,53	0,54	0,53	0,53
	C3	0,62	0,66	0,67	0,63	0,67	0,66
	C4	0,69	0,7	0,7	0,68	0,69	0,68
	Pz	0,54	0,6	0,61	0,52	0,59	0,58
6	Fz	0,94	0,92	0,93	0,93	0,93	0,94
	Cz	0,88	0,87	0,89	0,87	0,88	0,88
	C3	0,95	0,95	0,96	0,95	0,95	0,95
	C4	0,98	0,97	0,97	0,97	0,97	0,98
	Pz	0,96	0,96	0,96	0,96	0,96	0,96
8	Fz	0,67	0,67	0,67	0,67	0,67	0,67
	Cz	0,59	0,62	0,64	0,56	0,63	0,62
	C3	0,65	0,65	0,65	0,64	0,65	0,66
	C4	0,72	0,73	0,74	0,72	0,74	0,74
	Pz	0,99	0,99	0,99	0,99	0,99	0,99
10	Fz	0,65	0,64	0,64	0,65	0,65	0,66
	Cz	0,75	0,75	0,74	0,74	0,76	0,76
	C3	0,65	0,64	0,64	0,64	0,66	0,65
	C4	0,79	0,78	0,79	0,8	0,8	0,8
	Pz	0,88	0,87	0,87	0,88	0,88	0,88

Ao comparar as acurácias dos classificadores, o SVM, MLP e Gradient Boosting tiveram as melhores performances na maioria dos sujeitos. O canal C4 foi superior para os sujeitos 1, 5 e 6, enquanto o canal Pz teve melhor desempenho nos outros sujeitos, com 99% de acurácia no sujeito 8. A Tabela 1 revela que o sujeito 6 teve as melhores acurácias gerais.

Conclusões

Após a análise dos resultados, é evidente que a detecção automatizada da sonolência utilizando técnicas de machine learning com dados de EEG tem implicações significativas para a segurança e eficiência operacional na indústria de óleo e gás. As operações offshore, especialmente nas salas de controle, exigem vigilância constante e decisões precisas para garantir a segurança dos trabalhadores e a proteção ambiental. Identificar e prevenir a sonolência entre os operadores pode reduzir o risco de acidentes e derramamentos de petróleo, salvaguardando vidas humanas e ecossistemas marinhos. Esta pesquisa não só abre novas perspectivas para aprimorar a segurança nas operações offshore, mas também destaca a importância de investir em tecnologias inovadoras para enfrentar os desafios únicos da indústria.

Explorando diversos algoritmos de machine learning, modelos como SVM, MLP e Gradient Boosting demonstraram altas acurácias na classificação dos estados de alerta e sonolência. A escolha dos canais de EEG, como C4 e Pz, impactou o desempenho do modelo, ressaltando a necessidade de uma seleção cuidadosa das características para cada indivíduo e cenário. O modelo de ensemble apresentou resultados comparáveis aos melhores modelos individuais, demonstrando sua eficácia e robustez na detecção de sonolência. A performance do Sujeito 6, com altas acurácias em canais como Fz, Cz, C3 e C4, destaca a importância de considerar variações individuais. Além disso, a utilização de múltiplos canais de EEG nem sempre resulta em melhor desempenho, sugerindo que a coleta extensiva de canais pode ser dispendiosa e incômoda. Futuras pesquisas devem explorar modelos de aprendizado profundo e considerar a inclusão de outros sinais biológicos para aprimorar os métodos de detecção de sonolência.

Agradecimentos

Os autores agradecem às seguintes agências brasileiras de fomento à pesquisa pelo apoio financeiro: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), 308924/2021-5, 310892/2022-8, 402761/2023-5; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Código de Financiamento 001; Programa de Recursos Humanos 38.1 (PRH 38.1) intitulado Análise de Risco e Modelagem Ambiental na Exploração, Desenvolvimento e Produção de Petróleo e Gás, financiado pela Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP) e gerenciado pela Financiadora de Estudos e Projetos (FINEP), 044819.

Bibliografia

- [1] S. A. M. Naqvi, M. Raza, S. Ghazal, S. Salehi, Z. Kang, and C. Teodoriu, “Simulation-based training to enhance process safety in offshore energy operations: Process tracing through eye-tracking,” *Process Safety and Environmental Protection*, vol. 138, pp. 220–235, Jun. 2020, doi: 10.1016/j.psep.2020.03.016.
- [2] F. E. A. Horita, J. P. de Albuquerque, and V. Marchezini, “Understanding the decision-making process in disaster risk monitoring and early-warning: A case study within a control room in Brazil,” *International Journal of Disaster Risk Reduction*, vol. 28, pp. 22–31, Jun. 2018, doi: 10.1016/j.ijdr.2018.01.034.
- [3] P. R. Schulman and E. Roe, “A control room metric for evaluating success and failure in high reliability crisis management,” *Policy Soc*, vol. 30, no. 2, pp. 129–136, May 2011, doi: 10.1016/j.polsoc.2011.03.007.
- [4] L. G. Militello, E. S. Patterson, L. Bowman, and R. Wears, “Information flow during crisis management: challenges to coordination in the emergency operations center,” *Cognition, Technology & Work*, vol. 9, no. 1, pp. 25–31, Feb. 2007, doi: 10.1007/s10111-006-0059-3.
- [5] J. Isimite and P. Rubini, “A dynamic HAZOP case study using the Texas City refinery explosion,” *J Loss Prev Process Ind*, vol. 40, pp. 496–501, Mar. 2016, doi: 10.1016/j.jlp.2016.01.025.
- [6] T. A. Saurin and S. S. Gonzalez, “Assessing the compatibility of the management of standardized procedures with the complexity of a sociotechnical system: Case study of a control room in an oil refinery,” *Appl Ergon*, vol. 44, no. 5, pp. 811–823, Sep. 2013, doi: 10.1016/j.apergo.2013.02.003.
- [7] F. Rundo *et al.*, “An Innovative Deep Learning Algorithm for Drowsiness Detection from EEG Signal,” *Computation*, vol. 7, no. 1, p. 13, Feb. 2019, doi: 10.3390/computation7010013.
- [8] I. Stancin, M. Cifrek, and A. Jovic, “A Review of EEG Signal Features and Their Application in Driver Drowsiness Detection Systems,” *Sensors*, vol. 21, no. 11, p. 3786, May 2021, doi: 10.3390/s21113786.
- [9] P. M. S. Ramos, C. B. S. Maior, M. C. Moura, and I. D. Lins, “Automatic drowsiness detection for safety-critical operations using ensemble models and EEG signals,” *Process Safety and Environmental Protection*, vol. 164, pp. 566–581, Aug. 2022, doi: 10.1016/j.psep.2022.06.039.
- [10] C. B. S. Maior, M. J. das C. Moura, J. M. M. Santana, and I. D. Lins, “Real-time classification for autonomous drowsiness detection using eye aspect ratio,” *Expert Syst Appl*, vol. 158, p. 113505, Nov. 2020, doi: 10.1016/j.eswa.2020.113505.
- [11] V. Bajaj, S. Taran, S. K. Khare, and A. Sengur, “Feature extraction method for classification of alertness and drowsiness states EEG signals,” *Applied Acoustics*, vol. 163, p. 107224, Jun. 2020, doi: 10.1016/j.apacoust.2020.107224.

- [12] B. Venkata Phanikrishna, A. Jaya Prakash, and C. Suchismitha, "Deep Review of Machine Learning Techniques on Detection of Drowsiness Using EEG Signal," *IETE J Res*, vol. 69, no. 6, pp. 3104–3119, Aug. 2023, doi: 10.1080/03772063.2021.1913070.
- [13] Q. Massoz, T. Langohr, C. Francois, and J. G. Verly, "The ULg multimodality drowsiness database (called DROZY) and examples of use," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, Mar. 2016, pp. 1–7. doi: 10.1109/WACV.2016.7477715.
- [14] L. Cabañero-Gomez, R. Hervas, I. Gonzalez, and L. Rodriguez-Benitez, "eeglib: A Python module for EEG feature extraction," *SoftwareX*, vol. 15, p. 100745, Jul. 2021, doi: 10.1016/j.softx.2021.100745.
- [15] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D*, vol. 31, no. 2, pp. 277–283, Jun. 1988, doi: 10.1016/0167-2789(88)90081-4.
- [16] V. P. B and S. Chinara, "Automatic classification methods for detecting drowsiness using wavelet packet transform extracted time-domain features from single-channel EEG signal," *J Neurosci Methods*, vol. 347, p. 108927, Jan. 2021, doi: 10.1016/j.jneumeth.2020.108927.
- [17] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalogr Clin Neurophysiol*, vol. 29, no. 3, pp. 306–310, Sep. 1970, doi: 10.1016/0013-4694(70)90143-4.
- [18] O. Kramer, "Scikit-Learn," 2016, pp. 45–53. doi: 10.1007/978-3-319-33383-0_5.
- [19] S. J. Rigatti, "Random Forest," *J Insur Med*, vol. 47, no. 1, pp. 31–39, Jan. 2017, doi: 10.17849/in-sm-47-01-31-39.1.
- [20] X. Wang, J. Wang, K. Zhang, F. Lin, and Q. Chang, "Convergence and objective functions of noise-injected multilayer perceptrons with hidden multipliers," *Neurocomputing*, vol. 452, pp. 796–812, Sep. 2021, doi: 10.1016/j.neucom.2020.03.119.
- [21] I. Belakhdar, W. Kaaniche, R. Djemal, and B. Ouni, "Single-channel-based automatic drowsiness detection architecture with a reduced number of EEG features," *Microprocess Microsyst*, vol. 58, pp. 13–23, Apr. 2018, doi: 10.1016/j.micpro.2018.02.004.
- [22] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mech Syst Signal Process*, vol. 21, no. 6, pp. 2560–2574, Aug. 2007, doi: 10.1016/j.ymssp.2006.12.007.
- [23] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update," *J Neural Eng*, vol. 15, no. 3, p. 031005, Jun. 2018, doi: 10.1088/1741-2552/aab2f2.
- [24] A. Dogan and D. Birant, "A Weighted Majority Voting Ensemble Approach for Classification," in *2019 4th International Conference on Computer Science and Engineering (UBMK)*, IEEE, Sep. 2019, pp. 1–6. doi: 10.1109/UBMK.2019.8907028.
- [25] E.-A. MINASTIREANU and G. MESNITA, "Light GBM Machine Learning Algorithm to Online Click Fraud Detection," *Journal of Information Assurance & Cybersecurity*, pp. 1–12, Apr. 2019, doi: 10.5171/2019.263928.
- [26] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors)," *The Annals of Statistics*, vol. 28, no. 2, Apr. 2000, doi: 10.1214/aos/1016218223.
- [27] Chin-Teng Lin, Ruei-Cheng Wu, Sheng-Fu Liang, Wen-Hung Chao, Yu-Jie Chen, and Tzyy-Ping Jung, "EEG-based drowsiness estimation for safety driving using independent component analysis," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 12, pp. 2726–2738, Dec. 2005, doi: 10.1109/TCSI.2005.857555.
- [28] S. Kumari, D. Kumar, and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 40–46, Jun. 2021, doi: 10.1016/j.ijcce.2021.01.001.
- [29] M. Saqlain, B. Jargalsaikhan, and J. Y. Lee, "A Voting Ensemble Classifier for Wafer Map Defect Patterns Identification in Semiconductor Manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, vol. 32, no. 2, pp. 171–182, May 2019, doi: 10.1109/TSM.2019.2904306.
- [30] Z.-H. Zhou, "Open-environment machine learning," *Natl Sci Rev*, vol. 9, no. 8, Aug. 2022, doi: 10.1093/nsr/nwac123.