Mental Workload Classification from fNIRS Signals by Leveraging Machine Learning

M. Hasan[1], M. Mahmud[2], S. Poudel[3], K. Donthula[4] and K. Poudel[1,3]

[1]Computational and Data Science, Middle Tennessee State University, Murfreesboro, TN, USA
[2]Department of Electrical and Computer Engineering, University of Memphis, TN, USA
[3]Computer Science, Middle Tennessee State University, Murfreesboro, TN, USA
[4]Senior Software Engineer, Optum Inc, Murfreesboro, TN, USA
mh2ay@mtmail.mtsu.edu, mmahmud@memphis.edu, sp2ai@mtmail.mtsu.edu,
kiran.donthula@optum.com, khem.poudel@mtsu.edu

***Abstract*— Mental workload (MWL) identification is vital to know human cognitive functioning, performance, and well-being. In this work, we develop models for identifying low vs. high MWL using different genres of machine learning classifiers. We used non-invasive functional near-infrared spectroscopy (fNIRS) signals while participants classified the low vs. high levels of MWL tasks. Our analysis shows the low vs. high MWL can be identified best from the whole brain data. The k-nearest neighbors classifier showed the best performance with an accuracy of 98.8%, an area under the curve (AUC) of 98.8%, F1 score, precision, and recall of 98.0% from the whole brain data without overlapping signals. A separate hemisphere analysis using left hemisphere (LH) and right hemisphere (RH) activity showed that the LH activity has better classification ability than the RH activity. We also examined the classification with the top six features that could identify the low vs. high MWL with an accuracy of 97.4%, (AUC) 97.4%, F1 score, precision, and recall of 97.0%. These findings would be useful for developing more intuitive and user-friendly interfaces in the human-computer interface.**

Keywords*— *Functional near-infrared spectroscopy (fNIRS); Machine learning; Feature selection; KNN; and XGBoost.

I. Introduction

Mental workload (MWL) describes the cognitive demand that is placed on an individual during a specific task. The study of MWL is paramount for understanding how the human brain processes information, ergonomics, and copes with different cognitive demands [1]. Researchers demonstrated that higher MWL leads to lower task performance even in known knowledge-based task activity [2, 3].

Some researchers [4, 5] examined the MWL classification from the electroencephalogram (EEG), combined EEG and heart rate, functional near-infrared spectroscopy (fNIRS), and combined fNIRS & EEG. However, the fNIRS modality offers portability, fewer motion artifacts, and high spatial resolution, making it attractive in real-world applications. The fNIRS captures the concentration changes of oxygenated hemoglobin and deoxygenated hemoglobin in the brain's cortical areas. These measurements provide valuable insights for understanding cognitive processing. Researchers [6, 7] demonstrated that MWL can be classified from the fNIRS-derived prefrontal activity using a convolution neural network. Rojas et al. [8] identified the biomarker of human pain using machine learning (ML) from the fNIRS signals. We are inspired by the previous [5, 9, 10] research on fNIRS-based neurorehabilitation, neurodevelopmental studies, and human-computer interface that provides promising results.

A handful of research was conducted on hemisphere analysis on MWL classification. Huang et al. examined that during the MWL classification, more connectivity was created in the left hemisphere (LH) as compared to the right hemisphere (RH) via EEG-based analysis [11]. Ma et al. [12] investigated using functional magnetic resonance imaging (fMRI) and found that higher MWL enhances the connectivity of the LH. However, none could explore MWL identification using individual hemisphere data from the fNIRS modality. We investigated under the test battery set whether LH dominates during the classification of MWL.

Perhaps, the concentration changes of blood oxygenation and deoxygenation are not the same during the high and low MWL task identification. The intensity and phase could differ in the LH and RH of the brain. Those measurements could be useful to identify the high vs. low MWL situation. Researchers used overlapping window basis analysis to identify MWL and found decent classification results [13]. However, they could not explore without overlapping signals. Departing from the previous sliding window overlapping approach and deep learning-based MWL classification [11]. In this current work, we proposed that the non-overlapping fNIRS signal can distinguish the low vs. high MWL by leveraging classical ML models.

Our main contributions are as follows:

- We investigated that the MWL can be classified well using the KNN classifier from fNIRS signals.
- We also identified the top six important features that could predict the low vs. high MWL as slightly lower than the using full-brain features.
- Furthermore, we examined the individual hemisphere analysis that shows that LH is dominant in the MWL classification task as compared to RH.

The rest of the paper is organized as follows. We describe the dataset and ML models used in this study in Section II. Subsequently, Section III discusses the results and concludes the paper with key findings in Section IV.

II. Materials & Methods

Classification of MWL from the fNIRS-derived signals associated dataset with this study is reported in detail [13]. This data was originally collected and designed by the two collaborating teams in the Department of Computer Science and Biomedical Engineering at Tufts University, MA, USA [13]. In the current study, we present a new analysis of how well the functional brain activity (i.e., fNIRS) can classify the low vs. high MWL and dominance in the hemisphere. In addition, how well the selected important features classify the MWL as compared to all features.

II-A. Participants

There are 68 participants in this study (32 Asian, 27 White, 3 Black, 2 Hispanic, 1-Pacific Islander, and 3 other races; aged 18 to 44 years). Participants were recruited from around the Tufts University campus in Medford, MA, USA. Each participant sat on a standard chair in front of a desktop computer screen, keyboard, and mouse for 30-60 minutes. All participants were English speakers, and none had reported any history of neurological disease. The 64 participants were right-handed, 3 left-handed, and 1 unknown [14]. Each participant received compensation for their time. The procedures for this study were approved by the Institutional Review Board at Tuft University. All participants gave informed written consent about data release for the public.

II-B. Stimuli & Task

Four n-back (i.e., 0-back, 1-back, 2-back, and 3-back) stimuli were presented in this experiment. The stimulus presentation scenario is depicted in Figure 1. Where 0-back is the lowest MWL, and 3-back is the highest MWL. The stimuli were presented at a Latino square flatted version of a 4 x 4 array with different 4 n-back tasks (e.g., n = 0, 1 , 2, 3). Each stimulus was presented for 0.5 second, then 1.5 seconds for hidden (total of 2 seconds) [13]. During the hidden period, participants were asked to label them by pressing the keyboards for the defined specific n-back task. If the digit was matched with target n-back, then press the left arrow and otherwise the right arrow. The system advanced to the next trial after 2 seconds, regardless of whether the subject pressed the button. Before starting the n-back task, an infographic instruction was given to remind the subject of the present value of n (e.g., 0, 1, 2, 3). Each n-back was presented for 40 trials. Subjects were rested for 20 seconds for 0 and 1 back block; 2 and 3 back blocks had longer rest periods of 30 and 40 seconds, respectively.

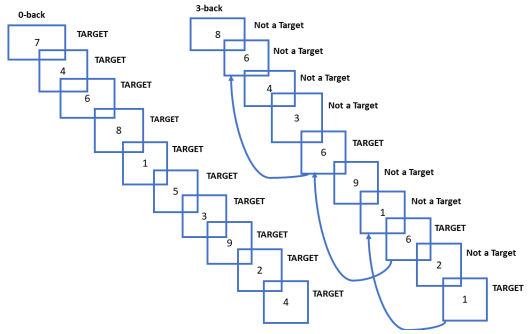

Figure 1. Stimulus presentation.

II-C. fNIRS Recordings and Data Preprocessing

An Imagent frequency-domain (FD) NIRS instrument manufactured by ISS (Champaign, IL, USA) was used for recording the intensity and phase of changes of the blood oxygenated and deoxygenated hemoglobin [13]. During the behavioral task classification, the fNIRS signal was recorded via a two probes headband from the left and right sides of the forehead. Raw FD-NIRS measurements contain the trace of alternating current intensity and phase changes. Ultimately, these measurements provide changes in oxyhemoglobin and deoxyhemoglobin concentration from either phase or intensity. During this experiment, the DC voltage was monitored to maintain the voltage within the range for good signal quality. Over-voltage and saturation can shut down the fNIRS systems automatically, and the users feel discomfort. The recorded data was sampled at 5.2 Hz and filtered (0.001-0.2 Hz) to remove the artifacts (e.g., drift, respiration, and heartbeat issues)[13].

II-D. Features Extractions

The preprocessed fNIRS recording provided eight measurements from two hemispheres (LH and RH) and considered them as features. Each hemisphere has four measurements (i.e., two intensities for oxygenated and deoxygenated; two phases change for oxygenated and deoxygenated). We used them as a feature vector. The grand averaged of eight features is represented in Figure 2. The grand average of the functional brain activity for 0-back and 3-back are not the same values. For high MWL, the grand average produced higher activity compared to the low MWL. We conducted an independent 2-sample t-test between 0-back and 3-back of each feature (i.e., brain activity). We found that all the features showed statistical significance (p-value <0.0005). We assume these features could be useful for identifying low vs. high MWL.

II-E. Classification

We used different genres of ML classifiers to examine the low vs. high MWL classification. K-nearest neighbors (KNN), decision trees (DT), extreme gradi-

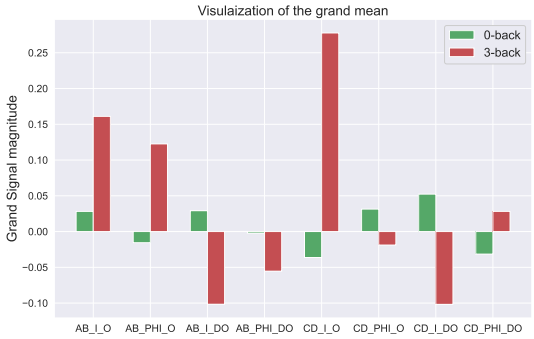

Figure 2. Grand average of each feature measurement while classifying low (i.e., 0-back) and high (i.e., 3-back) MWL tasks. AB: left hemisphere; CD: right hemisphere; I: Intensity (µmol/L); PHI: phase (rad); O: oxyhemoglobin; DO: deoxy-hemoglobin.

ent boosting (XGBoost), light gradient boosting (light-GBM), random forest (RF), and support vector machine (SVM) are widely employed as robust classifiers for detecting subtle patterns, handling intricate datasets, and performing classification tasks in diverse domains, including the field of neuroscience, computer vision, and text classification. Classifiers performance significantly affects by the selection of their hyperparameters. To identify the optimal parameters, we applied a grid search approach for fine-tuning the hyperparameters for each classifier [15]. The grid search process provided the best parameters. Then we trained the model on the training data set with optimal hyperparameters. Once the model was trained, then we provided the features only from the test dataset and predicted the low vs. high MWL. The test dataset models were never seen. The classifier performance metrics (accuracy, precision, recall, and F1 scores) were calculated from the standard formula [16]. The area under the curve (AUC) metric provides valuable insight into the ability of a classifier and how well it distinguishes classes. An excellent model that has good separability, the AUC value is closer to 1. On the other hand, in a poor-performing model that has no ability to distinguish classes, the AUC is closer to 0.

For all classifiers, the data were z-score normalized (i.e., zero mean and one standard deviation) to ensure all features were on a common scale range [17]. Then we split the data training and test set 80% and 20% respectively [18].

Hyperparameter Optimization: Hyperparameter optimization is crucial to build an effective classification framework. Hyperparameters significantly influence the model performance and generalization abilities. We fine-tuned the parameters with the grid search approach [24], [25] to obtain the optimal parameters. These optimal parameters provide a generalization of the model and better prediction ability. For each ML classifier, we selected a range of model parameters with five-fold cross-validation [19] for identifying the optimal hyperparameters and used them in our final model.

K-Nearest Neighbors Algorithm: KNN is a popular machine learning model that is used for regression and classification problems. This can be used in a supervised and unsupervised way. When a target is known along with features, then it can be used as supervised, otherwise unsupervised. There are several parameters of the KNN classifier. The optimal values of the parameters reduce the bias and variance and make the model generalize. The KNN classifier is used for healthcare big data classification [20] anomaly intrusion detection [21], facial expression recognition [22], and found excellent results. In this work, we set the optimal parameters: metric =‘minkowski’, n_neighbors = 2, weights = ‘distance’, p = 2 that achieved from the grid search technique. *Decision Tree (DT)*: DT is a non-parametric supervised ML algorithm that uses for classification and regression analysis. This algorithm learns by recursively partitioning the feature space into regions based on the feature values and creating a tree-like structure. DT can handle both numerical and categorical data without requiring extensive preprocessing [23]. The predicted class for new data points is identified from the node to a leaf. The node represents a decision based on the features, and the node leaf represents the final decisions. It is a highly interpretable model [24]. In this work, we selected the optimal parameters: criterion = ‘entropy’,‘splitter = ‘best’, and min_samples_split = 2 that were yielded by the grid search approach.

XGBoost (Extreme Gradient Boosting): It is widely used for classification and regression analysis. XGBoost uses gradient boosting algorithms that provide minimum loss function [25, 26]. The XGBoost model can provide the features importance. This classifier supports parallel processing and makes it scalable and faster for larger datasets. Its main benefits are automatically learn how to handle missing values during training phase and high predictive accuracy. During the training phase, we set the hyper parameters: objective =‘binary : logistic’, random_state = 42, min_child_weight = 1, max_depth = 15, learning_rate= 0.3, gamma = 0.2, colsample_bytree= 0.7, eval_metric= ‘logloss’.

Light Gradient Boosting (LightGBM):LightGBM is based on a gradient-boosting algorithm with ensemble learning techniques used for regression and classification tasks. This framework has the advantages of faster training speed, high performance, lower memory usage, support for parallel learning, and the capability to handle large-scale datasets. It is used in many fields, including computer vision, natural language processing, financial analysis, and healthcare classification [27]. We set the optimal hyperparameters: colsample_bytree = 0.7, learning_rate = 0.3, max_depth = 100, min_child_weight = 0.5, n_estimators = 200.

Random Forest (RF): RF classifier is an ensemble learning method that consists of a tree-based classifier [28]. This algorithm is used for classification, regression, and feature selection. The advantage of this classifier is being capable of handling large-scale data. This classifier is used in healthcare, text classification, and many other real word applications. We trained the RF model with the optimal hyperparameters max_depth = 40, ax_features = 'log2', n_estimators = 200 that were found by the grid search approach.

Support Vector Machine (SVM): SVM is remarkably used in classification and regression analysis due to handling high dimensional data with generalize well on unseen examples [29, 30]. It provides maximum margin between classes by selecting the optimal hyperplane that best separates the data into distinct classes. SVM seeks the optimal decision boundary that maximizes the margin and provides less error. The key advantage of SVM is the ability to handle linear and non-linearly separable data by using kernel trick. During the training phase, we used the optimal parameters: C = 10, kernel = 'rbf', degree = 3, and, $\gamma = 0.1$ that were obtained by the grid search approach.

III. Results

III-A. *Classification of Low vs. High Mental Workload from Whole Brain and Hemisphere Data*

We first investigated how the low vs. high MWL could be classified from the whole brain data and then hemisphere individuals (LH and RH) data. We represent the six classifiers' accuracy in Figure 3 for whole-brain and Figure 4 for hemisphere (i.e., LH and RH). The accuracy and other performance metrics are reported in Table 1. Using whole brain data, the KNN provided the best classification accuracy of 98.8%, with AUC of 98.8%, precision, recall, and F1-score of 99.0%. The RF, DT, XGBoost, and lightGBM showed a slightly lower accuracy as compared to KNN. However, SVM exhibited the lowest classification performance (accuracy 69.5%, with AUC 69.5%, precision, recall, and F1-score 70.0%).

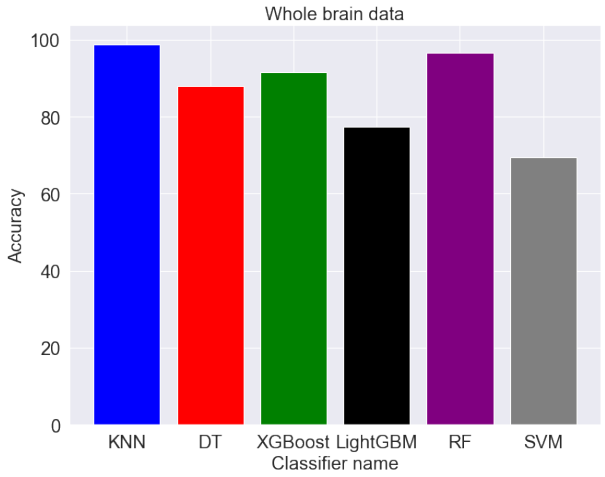

Figure 3. Shows the classification of low vs. high MWL using full brain data.

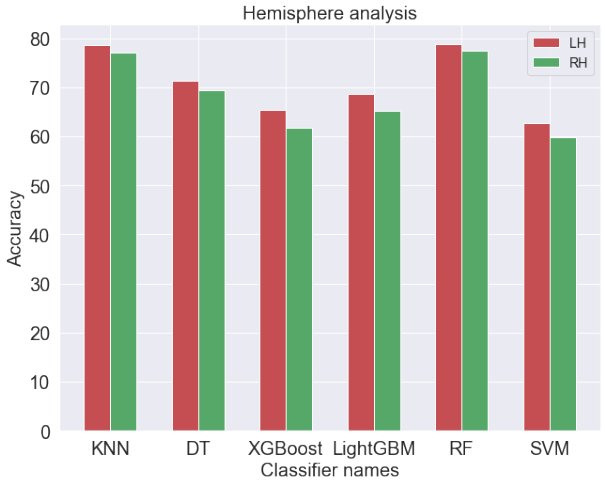

Figure 4. Shows the classification of low vs. high MWL using hemisphere data.

Classification using hemispheres data showed a lower performance as compared to the whole brain data. All classifiers showed that the LH data yielded better MWL classification than RH data. Interestingly, in the hemispheres data, the RF classifier provided the best classification accuracy, followed by KNN, XGBoost, then light GBM. The RF classifier yielded accuracy of 78.8% with AUC 78.8% using LH data where as on RH accuracy of 77.5% and AUC 77.5%. However, the SVM exhibited the lowest performance (LH: accuracy 62.7% and AUC 62.7% ; RH: accuracy 59.9% and AUC 59.0%) among all. Our hemisphere results corroborate with previous studies of LH dominating in MWL [11, 12].

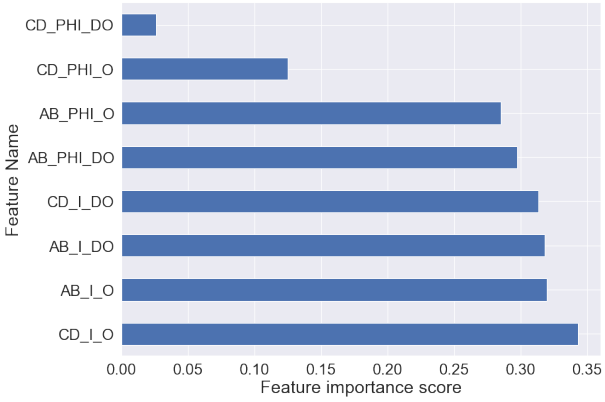

Figure 5. Visualization of features with their importance ranked (by permutation feature selection). y-axis represents features name; x-axis: represents feature score; otherwise for alphabet notation Figure 2.

Table 1. KNN, DT, XGBoost, LigghtGBM, RF, and SVM classifiers' performance metrics (%) for detecting low vs. high MWL.

Classifiers name	Average measure(%)	Whole-brain data	LH's data	RH's data
KNN	Accuracy	98.8	78.5	77.1
	AUC	98.8	78.5	77.1
	Precision	99.0	78.0	77.0
	Recall	99.0	78.0	77.0
	F1-score	99.0	78.0	77.0
DT	Accuracy	87.9	71.4	69.4
	AUC	87.9	71.4	69.4
	Precision	88.0	71.0	69.0
	Recall	88.0	71.0	69.0
	F1-score	88.0	71.0	69.0
XGBoost	Accuracy	91.6	65.3	61.7
	AUC	91.6	65.3	61.7
	Precision	92.0	65.0	62.0
	Recall	92.0	65.0	62.0
	F1-score	92.0	65.0	62.0
LighGBM	Accuracy	77.3	68.7	65.2
	AUC	77.3	68.7	65.2
	Precision	77.0	69.0	65.0
	Recall	77.0	69.0	65.0
	F1-score	77.0	69.0	65.0
RF	Accuracy	96.7	78.8	77.5
	AUC	96.7	78.8	77.5
	Precision	97.0	79.0	78.0
	Recall	97.0	79.0	78.0
	F1-score	97.0	79.0	78.0
SVM	Accuracy	69.5	62.7	59.9
	AUC	69.5	62.7	59.9
	Precision	70.0	63.0	60.0
	Recall	69.0	63.0	60.0
	F1-score	69.0	63.0	60.0

III-B. Classification of Low vs. High Mental Workload from Top Six Features

We used permutation feature selection [31, 32] with the KNN classifier. We ranked the features according to the feature important scores that are represented in the horizontal bar chart in Figure 5. We selected the top six important features out of eight features and again trained the classifiers and examined the model performances on test dataset. We found that all classifiers exhibited similar whole brain features classification results but a slightly less by 2-5% performance degraded. The classification of low vs. high MWL obtained by six features is represented in Figure 6. With the six important features KNN classifier yielded (accuracy 97.4%, AUC 97.4%, precision, recall, and F1-score 97.0%). Followed by RF (accuracy 93.3%, AUC 93.3%, precision, recall, and F1-score 93.0%), XGBoost (accuracy 85.6%, AUC 85.6%, precision, recall, and F1-score 86.0%). The lowest classification accuracy was found by SVM classifier (accuracy 67.1%, AUC 67.1%, precision, recall, and F1-score 67.0%).

IV. Conclusions

Our results demonstrate that the MWL can be classified well from the fNIRS-derived non overlapping signals. Moreover, the LH data showed better MWL classification accuracy as compared to RH. Six critical features could classify the low vs. high MWL by reducing the accuracy of 2-5% as compared to whole brain data. This paper sheds light on the potential for enhancing performance in sports, driving, and healthcare sectors through a better understanding of MWL. Our analysis can be used in human-computer interfaces, transportation, and healthcare for detecting high MWL. This can lead to the design of user-friendly interfaces, automation, and stress-reduced application to enhance safety and performance. We need further analysis with multiple sensors fNIRS headset.

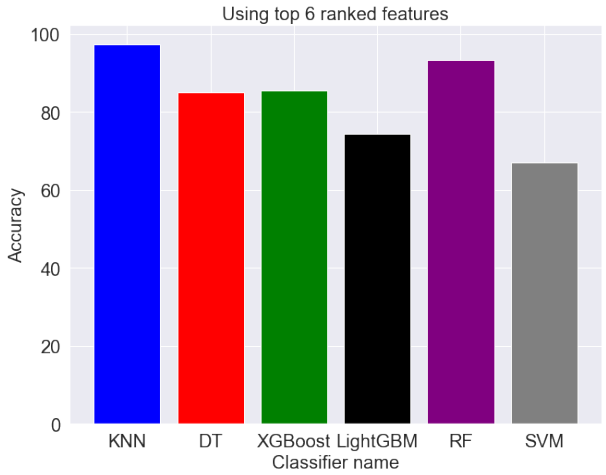

Figure 6. Classification of low vs. high MWL based on the top six ranked features.

Acknowledgments

This work was supported by the Computational Data Science program and the Department of Computer Science at the Middle Tennessee State University.

References

[1] H. Gu, H. Chen, Q. Yao, W. He, S. Wang, C. Yang, J. Li, H. Liu, X. Li, X. Zhao *et al.*, "Efficiency of the brain network is associated with the mental workload with developed mental schema," *Brain Sciences*, vol. 13, no. 3, p. 373, 2023.

[2] M. S. Young, K. A. Brookhuis, C. D. Wickens, and P. A. Hancock, "State of science: mental workload in ergonomics," *Ergonomics*, vol. 58, no. 1, pp. 1–17, 2015.

[3] E. Zahmat Doost and W. Zhang, "Mental workload variations during different cognitive office tasks with social media interruptions," *Ergonomics*, vol. 66, no. 5, pp. 592–608, 2023.

[4] U. Asgher, K. Khalil, Y. Ayaz, R. Ahmad, and M. J. Khan, "Classification of mental workload (mwl) using support vector machines (svm) and convolutional neural networks (cnn)," *2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*. IEEE, 2020, pp. 1–6.

[5] J. Cao, E. M. Garro, and Y. Zhao, "Eeg/fnirs based workload classification using functional brain connectivity and machine learning," *Sensors*, vol. 22, no. 19, p. 7623, 2022.

[6] V. K. Dhulipalla and M. A. A. H. Khan, "Mental workload classification from non-invasive fnirs signals through deep convolutional neural network," *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, 2022, pp. 1450–1455.

[7] J.-H. Park, "Mental workload classification using convolutional neural networks based on fnirs-derived prefrontal activity," *Available at SSRN 4493010*.

[8] R. Fernandez Rojas, X. Huang, and K.-L. Ou, "A machine learning approach for the identification of a biomarker of human pain using fnirs," *Scientific reports*, vol. 9, no. 1, p. 5645, 2019.

[9] J. Benerradi, H. A. Maior, A. Marinescu, J. Clos, and M. L. Wilson, "Exploring machine learning approaches for classifying mental workload using fnirs data from hci tasks," *Proceedings of the Halfway to the Future Symposium 2019*, 2019, pp. 1–11.

[10] F. Putze, C. Herff, C. Tremmel, T. Schultz, and D. J. Krusienski, "Decoding mental workload in virtual environments: a fnirs study using an immersive n-back task," *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 3103–3106.

[11] J. Huang, Z. Traylor, S. Choo, and C. S. Nam, "Neural correlates of mental workload during multitasking: a dynamic causal modeling study," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 65, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2021, pp. 1337–1341.

[12] L. Ma, J. L. Steinberg, K. M. Hasan, P. A. Narayana, L. A. Kramer, and F. G. Moeller, "Working memory load modulation of parieto-frontal connections: Evidence from dynamic causal modeling," *Human brain mapping*, vol. 33, no. 8, pp. 1850–1867, 2012.

[13] Z. Huang, L. Wang, G. Blaney, C. Slaughter, D. McKeon, Z. Zhou, R. Jacob, and M. C. Hughes, "The tufts fnirs mental workload dataset & benchmark for brain-computer interfaces that generalize," *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.

[14] R. C. Oldfield, "The assessment and analysis of handedness: the edinburgh inventory," *Neuropsychologia*, vol. 9, no. 1, pp. 97–113, 1971.

[15] S. Poudel, R. Paudyal, B. Cankaya, N. Sterlingsdottir, M. Murphy, S. Pandey, J. Vargas, and K. Poudel, "Correction to: Cryptocurrency price and volatility predictions with machine learning," *Journal of Marketing Analytics*, pp. 1–1, 09 2023.

[16] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets," *PloS one*, vol. 10, no. 3, p. e0118432, 2015.

[17] H. A. Prihanditya *et al.*, "The implementation of z-score normalization and boosting techniques to increase accuracy of c4. 5 algorithm in diagnosing chronic kidney disease," *Journal of Soft Computing Exploration*, vol. 1, no. 1, pp. 63–69, 2020.

[18] M. S. Mahmud, F. Ahmed, M. Yeasin, C. Alain, and G. M. Bidelman, "Multivariate models for decoding hearing impairment using eeg gamma-band power spectral density," *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–7.

[19] T.-T. Wong and P.-Y. Yeh, "Reliable accuracy estimates from k-fold cross validation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 8, pp. 1586–1594, 2019.

[20] W. Xing and Y. Bei, "Medical health big data classification based on knn classification algorithm," *IEEE Access*, vol. 8, pp. 28 808–28 819, 2019.

[21] L. Kuang and M. Zulkernine, "An anomaly intrusion detection method using the csi-knn algorithm," *Proceedings of the 2008 ACM symposium on Applied computing*, 2008, pp. 921–926.

[22] K. S. Yadav and J. Singha, "Facial expression recognition using modified viola-john's algorithm and knn classifier," *Multimedia Tools and Applications*, vol. 79, no. 19-20, pp. 13 089–13 107, 2020.

[23] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, 2021.

[24] A. Priyam, G. R. Abhijeeta, A. Rathee, and S. Srivastava, "Comparative analysis of decision tree classification algorithms," *International Journal of current engineering and technology*, vol. 3, no. 2, pp. 334–337, 2013.

[25] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou *et al.*, "Xgboost: extreme gradient boosting," *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.

[26] C. Wade and K. Glynn, *Hands-On Gradient Boosting with XGBoost and scikit-learn: Perform accessible machine learning and extreme gradient boosting with Python*. Packt Publishing Ltd, 2020.

[27] D. D. Rufo, T. G. Debelee, A. Ibenthal, and W. G. Negera, "Diagnosis of diabetes mellitus using gradient boosting machine (lightgbm)," *Diagnostics*, vol. 11, no. 9, p. 1714, 2021.

[28] M. Pal, "Random forest classifier for remote sensing classification," *International journal of remote sensing*, vol. 26, no. 1, pp. 217–222, 2005.

[29] M. N. Hasan, S. Hamdan, S. Poudel, J. Vargas, and K. Poudel, "Prediction of length-of-stay at intensive care unit (icu) using machine learning based on mimic-iii database," *2023 IEEE Conference on Artificial Intelligence (CAI)*. IEEE, 2023, pp. 321–323.

[30] S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, and W. Xu, "Applications of support vector machine (svm) learning in cancer genomics," *Cancer genomics & proteomics*, vol. 15, no. 1, pp. 41–51, 2018.

[31] A. Altmann, L. Toloşi, O. Sander, and T. Lengauer, "Permutation importance: a corrected feature importance measure," *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, 2010.

[32] M. Ojala and G. C. Garriga, "Permutation tests for studying classifier performance." *Journal of machine learning research*, vol. 11, no. 6, 2010.