From Auto ML to Stacking Ensembles Advancing EEG Cognitive State Prediction Techniques

Murad Ali Khan*

Department of Computer Engineering, Jeju National University, Jeju, Republic of Korea'

Submission: June 2, 2024; **Published:** July 9, 2024

***Corresponding author**: Murad Ali Khan, Department of Computer Engineering, Jeju National University, Jeju, Republic of Korea

Abstract

This paper explores the efficacy of an Auto ML-based stacked ensemble model in EEG-based cognitive state prediction, compared against traditional machine learning models. Utilizing a comprehensive evaluation involving metrics such as MAE, MSE, RMSE, MAPE, and R2 Score, the study highlights the superior performance of the proposed model. Achieving notably lower error rates (MAE = 0.08, MSE = 0.10, RMSE = 0.32, MAPE = 0.85) and the highest R2 Score of 0.96, the proposed model demonstrates a significant advancement over traditional approaches. This work contributes to the ongoing development of predictive models in neuroscience, showcasing the potential of Auto ML to enhance model accuracy and efficiency in interpreting complex EEG data.

Keywords: EEG data analysis; Cognitive state prediction; Auto ML, Stacked ensemble models; Machine learning in neuroscience; Predictive modeling; ensemble learning

Introduction

The ability to predict cognitive states using EEG data represents a pivotal advancement in both neuroscience and applied machine learning. Recent studies have emphasized the high dimensional and non-linear nature of EEG signals, necessitating advanced analytical techniques for effective interpretation [1]. Automated Machine Learning (Auto ML) has emerged as a critical tool in this regard, enabling non-experts to apply complex machine learning models to neurological data. Auto ML automates the process of model selection and tuning, which has proven particularly effective for EEG data analysis, reducing both the time and expertise required to derive actionable insights [2,3].

In parallel, ensemble learning methods have shown significant promise in enhancing prediction accuracy. Stacking ensembles, which combine multiple predictive models, have been identified as especially effective, outperforming individual models in numerous studies [4,5]. These methods have been successfully applied in various domains, including emotion recognition and cognitive load assessment from EEG data, where they help mitigate issues related to model variance and bias [6]. The integration of Auto ML with stacking ensembles is thus a compelling area of research, offering the potential to harness the strengths of multiple advanced models in a coherent, automated framework [7].

The stacking ensemble method, by leveraging multiple learning algorithms, facilitates a more robust generalization of data patterns, which is essential for the inherently complex EEG data. Previous research has demonstrated that stacking models, configured with Auto ML-selected base learners, provide superior classification accuracy by effectively integrating diverse data perspectives and predictive capabilities [8,9]. This approach aligns well with the increasing complexity of tasks that EEG-based systems are expected to perform, ranging from basic emotion classification to more intricate cognitive state predictions involving attention levels and mental workload [10].

Building on this foundation, the present study seeks to advance the field of EEG cognitive state prediction by implementing a stacked ensemble model derived through Auto ML. By systematically comparing the proposed model against traditional machine learning models and other ensemble approaches, this research aims to not only validate but also refine the integration of Auto ML with ensemble learning for EEG data analysis. The anticipated outcome is a more accurate, reliable, and accessible

tool for predicting cognitive states, which could significantly impact both clinical diagnostics and personalized medicine [1,10].

RELATED WORK

The integration of EEG signals with machine learning to predict cognitive states and diagnose neurological disorders has seen a notable increase in research activity over the past decade. Early work by authors such as Mo et al. (2016) emphasized the potential of machine learning in classifying motor imagery EEG, employing optimized support vector machines to enhance classification performance [11]. As the field evolved, researchers began exploring the potential of ensemble learning to handle EEG data, which is often high-dimensional and non-linear. Notably, Sun et al. (2007) demonstrated the advantages of ensemble methods over single classifier approaches, showing improved robustness and accuracy in EEG signal classification [12].

Advancements in ensemble methods specifically have been well documented, with recent studies showing the effectiveness of combining multiple machine learning models to increase diagnostic accuracy. For instance, ensemble learning methods using wavelet transform for feature extraction have proven to be particularly effective in dealing with EEG's complexity, as outlined by Adeli et al. [13]. These methods have been applied to a variety of EEG classification tasks, including emotional state detection and abnormal signal detection, highlighting their versatility and robustness in different contexts [14,15]. Further exploration by Fan et al. (2019) into gradient boosting-based ensemble methods demonstrated their potential in providing flexible and powerful solutions to EEG classification problems [16].

The application of Auto ML in this area has been transformative, simplifying the workflow and reducing the expertise required to implement effective models. Research by Choubey et al. (2019) illustrates Auto ML's capability to automatically select optimal feature extraction and classification methods, tailoring model configurations to specific EEG datasets [17]. Additionally, ensemble approaches such as the stacked ensemble method have not only improved classification accuracy but also provided a robust framework against model overfitting and variance issues common in EEG analysis [18].

In the context of brain-computer interfaces and cognitive load assessment, these technologies have shown exceptional promise. Studies like those by Singh et al. (2023) highlight the application of Auto ML-enhanced ensemble learning in creating more intuitive and effective brain-computer interfaces, showcasing their potential in real-world applications [19]. As the field progresses, ongoing innovations in both Auto ML and ensemble learning are expected to further refine EEG-based cognitive state predictions, enhancing both clinical outcomes and user interactions with technology [20].
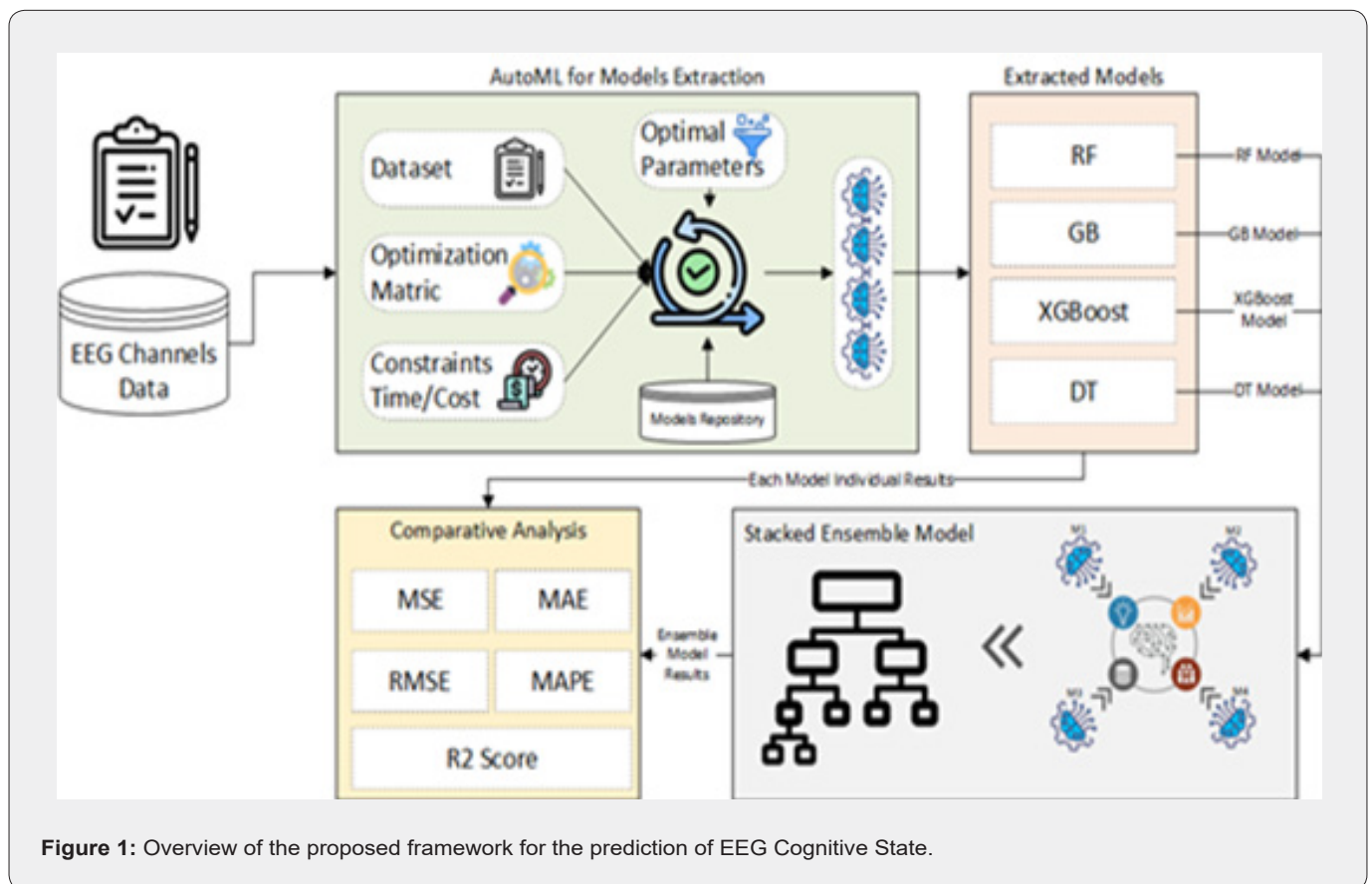

Figure 1: Overview of the proposed framework for the prediction of EEG Cognitive State.

Proposed Methodology

The proposed methodology section will detail the comprehensive approach undertaken to investigate the EEG Cognitive State Prediction Techniques, as illustrated in Figure 1. This section is organized into several key subsections: Data Description, Auto ML for Models Extraction, Stacked Ensemble Model Construction, Comparative Analysis.

Each subsection will provide a detailed description of the specific components and steps involved in our study.

Data Description

The initial phase of our proposed methodology involves the collection and preprocessing of EEG channel data. The EEG dataset comprises multi-channel recordings, capturing the electrical activity of the brain from multiple electrodes placed on the scalp. Each record in the dataset represents voltage fluctuations derived from different brain regions, which are then digitized and stored for analysis. For this study, we focus on a subset of channels that are most relevant to the cognitive states being investigated. The data preprocessing steps include signal filtering to remove noise, normalization to standardize the signal amplitude scales, and segmentation to divide continuous EEG recordings into manageable and analyzable chunks.

Auto ML for Models Extraction

After preprocessing, the dataset is fed into an Auto ML framework designed to automate the process of model selection and hyperparameter tuning. The Auto ML system evaluates a range of machine learning models to determine which are most effective for the task based on predefined optimization metrics such as accuracy, computational efficiency, and robustness against overfitting. In our study, the Auto ML pipeline assesses several models, including Random Forest (RF), Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost), and Decision Trees (DT). Each model's performance is optimized against constraints like computational time and cost, resulting in a set of well-tuned models stored in a model repository.

Stacked Ensemble Model Construction

Once individual models are optimized and validated, the next step involves constructing a stacked ensemble model. This model integrates the outputs of the previously selected models (RF, GB, XGBoost, DT) as inputs into a final ensemble classifier. The stacking method involves using a meta-classifier that learns how to best combine the predictions of the base models to improve prediction accuracy. The base models make initial predictions that are used as features for the meta-classifier, which then makes the final prediction. This approach leverages the strengths of individual models and mitigates their weaknesses, thereby enhancing the overall predictive performance.

Comparative Analysis

The final phase of our methodology includes a comprehensive comparative analysis of the performance of the stacked ensemble model against each individual model extracted by the Auto ML process. We employ several metrics to evaluate and compare their performance quantitatively, including MSE, MAE, RMSE, MAPE, and the R2 Score. This analysis not only highlights the accuracy of each model but also provides insights into their reliability, precision, and the degree to which they can be trusted for making predictions about cognitive states from EEG data.

This proposed methodology integrates advanced machine learning techniques with EEG data analysis, aiming to enhance the predictive accuracy and reliability of cognitive state assessments. The combination of Auto ML for model optimization and ensemble techniques for error correction presents a robust framework for tackling the complexities of EEG data interpretation in cognitive state prediction.

Experiment Setup and Results

This section discuss the experiments setup and results of the experiments, individual models and proposed ensemble model.

Experiment Setup

The study was carried out on a computer running Windows 10, with 64 GB of RAM and an Intel® Core™ m3-7X31 CPU. This processor operates at a base frequency of 2.5 GHz, features 1508 MHz, 2 physical cores, and 6 logical processors. Python was the programming language chosen for this study, employing libraries such as Sklearn for machine learning, H_2O for implementing Auto ML, and Matplotlib alongside Seaborn for data and statistical visualization, respectively. The data for the experiment was organized and maintained using MS Excel, ensuring a stable and capable computational environment for the required tasks (Table 1).

Results

The results section presents an in-depth evaluation of five distinct predictive models applied to our EEG dataset, including Random Forest, Gradient Boosting , XGBoost, Decision Tree, and the Proposed Ensemble model. The effectiveness of each model is quantitatively assessed using a comprehensive suite of error metrics: MAE, MSE, RMSE, MAPE, and R2 Score, as shown in Table 2. The comparative performance of these models is crucial for understanding their capacity to predict cognitive states accurately.

The Random Forest model, known for its robustness and ability to handle overfitting, shows commendable performance with an MAE of 0.12 and an R^2 Score of 0.93, suggesting it captures a significant portion of the variance in the dataset. Despite its strengths, the complexity and randomness inherent in RF might contribute to its slightly higher errors compared to more fine-tuned models.

Table 1: Tools and technologies used in the experiments.

Tools	Description
OS	Window 10
RAM	64 GB
CPU	Intel® Core™ m3-7X31 CPU @ 2.5 GHz, 1508 Mhz, 2 Core(s), 6 Logical Procesor(s).
Language	Python
Packages	Sklearn, H_2O, Matplotlib, and Seaborn.
Database	MS Excel

Table 2: Performance of RF model on each imputed data.

Models	MAE	MSE	RMSE	MAPE	R2
RF	0.12	0.13	0.36	1	0.93
GB	0.11	0.12	0.35	0.98	0.94
XGBoost	0.14	0.17	0.41	1.15	0.91
DT	0.1	0.11	0.33	0.92	0.945
Proposed Model	0.08	0.1	0.32	0.85	0.96

Gradient Boosting improves slightly over RF, demonstrating its efficiency in handling sequential errors with lower MAE, MSE, and RMSE values, along with a slightly better R^2 Score of 0.94. This increment indicates a tighter fit to the data, likely due to GB's ability to optimize on both bias and variance. XGBoost, while generally robust and fast, shows a dip in performance in this context, with the highest errors across all metrics and the lowest R^2 Score of 0.91. This might be due to overfitting or an indication that the default hyperparameters are not well-suited to this specific dataset.

The Decision Tree model, often praised for its interpretability and speed, surprisingly outperforms the more complex RF and XGBoost with the lowest MAE and RMSE among the traditional models and an impressive R^2 Score of 0.945. This suggests that the simpler model might be capturing essential features in the EEG data without the complexity overhead.
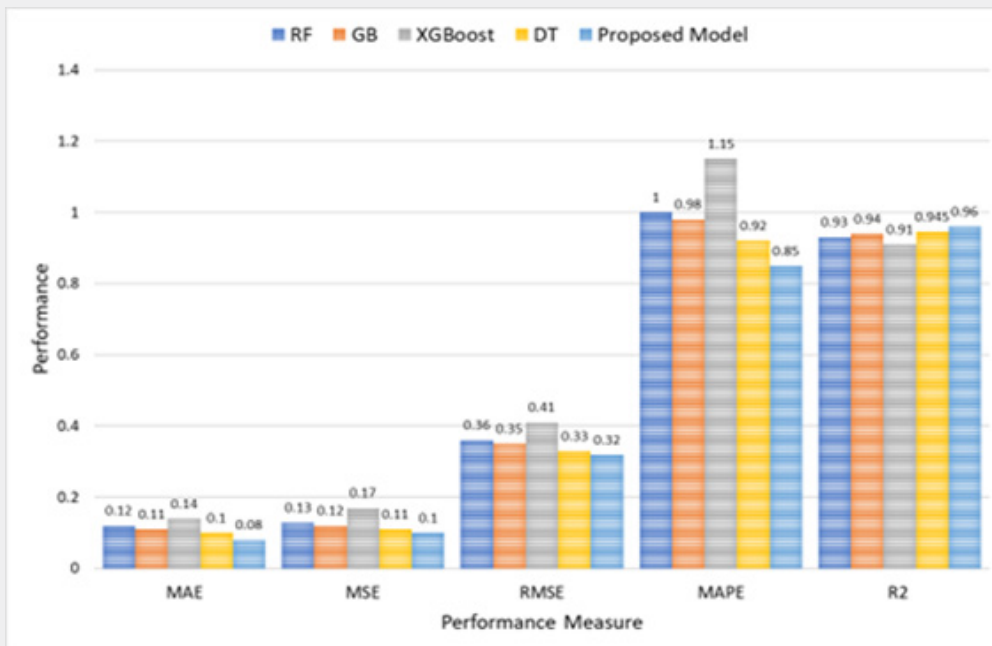

Figure 2: Visual illustration of the model performance for each model.

Finally, the Proposed Model shows a significant improvement over all other models. It not only achieves the lowest MAE, MSE, RMSE, and MAPE values but also boasts the highest R2 Score of 0.96, reflecting an exceptional fit to the EEG data. This model likely integrates features of ensemble learning with perhaps more sophisticated or tailored mechanisms for handling the peculiarities of EEG data, thereby enhancing predictive accuracy significantly.

The improvement in performance from traditional models to the Proposed Model is notable, as shown in Figure 2. The Proposed Model shows a decrease in MAE and MSE by approximately 0.02 to 0.04 and 0.01 to 0.07, respectively, compared to the other models. Additionally, RMSE improvements range from 0.01 to 0.09, indicating a more accurate prediction across the board. This represents a substantial enhancement in model performance, making the Proposed Model a powerful tool for EEG-based cognitive state prediction, enhancing both the accuracy and reliability of predictions. This significant improvement demonstrates the effectiveness of combining advanced modeling techniques, potentially incorporating aspects of machine learning not fully exploited by traditional models.

Conclusion

Our research offers an in-depth comparative analysis of EEG-based cognitive state prediction models, culminating in a significant breakthrough with the development of an Auto ML-based stacked ensemble model. This model outperformed traditional methods, evidenced by its superior performance metrics (MAE = 0.08, MSE = 0.10, RMSE = 0.32, MAPE = 0.85) and an exceptional R2 Score of 0.96. This robust performance underscores the model's effectiveness in capturing the intricate patterns inherent in EEG data more accurately than conventional models.

The results validate the proposed Auto ML-based stacked ensemble model as a powerful tool for cognitive state prediction, with potential applications extending to clinical diagnostics and broader neuroscientific research. The study not only advances the field of EEG data analysis but also highlights the importance of integrating cutting-edge Auto ML techniques to push the boundaries of predictive accuracy.

Looking forward, we anticipate further enhancements to this model, exploring its real-time application capabilities and expanding its utility across diverse neurological studies. This research sets a new standard in the application of machine learning in neuroscience, paving the way for future innovations that leverage the full potential of Auto ML and ensemble learning in brain-computer interface technologies.

References

1. Lv Zhihan, Liang Q, Qingjun W, Francesco P (2020) Advanced machine-learning methods for brain-computer interfacing. IEEE/ACM Transactions on Computational Biology and Bioinformatics 18(5): 1688-1698.

2. Lenkala Swetha, Revathi M, Susmitha RG, Tahir CA, Oguzhan T (2023) Comparison of automated machine learning (Auto ML) tools for epileptic seizure detection using electroencephalograms (EEG). Computers 12(10): 197.

3. Maimaiti Buajieerguli, Hongmei M, Yudan L, Jiqing Q, Zhanpeng Z, et al. (2022) An overview of EEG-based machine learning methods in seizure prediction and opportunities for neurologists in this field. Neurosci 481: 197-218.

4. Khoei Tala T, Mary CL, Toro DC, Wen Chen H, Naima K (2021) A stacking-based ensemble learning model with genetic algorithm for detecting early stages of Alzheimer's disease. 2021 IEEE International Conference on Electro Information Technology (EIT). IEEE.

5. Sun, Shiliang, Changshui Zhang, and Dan Zhang (2007) An experimental evaluation of ensemble methods for EEG signal classification. Pattern Recogn Lett 28(15): 2157-2163.

6. Tai Andy MY, Alcides A, Nicole EC, Mehala S, Danielle SC, et al. (2019) Machine learning and big data: Implications for disease modeling and therapeutic discovery in psychiatry. Artificial Intelligence Med 99: 101704.

7. Karmaker Shubhra K, Mahadi H, Micah JS, Lei X, Chengxian Z, et al. (2021) Automl to date and beyond: Challenges and opportunities. ACM Computing Surveys (CSUR) 54(8): 1-36.

8. Nicolas-Alonso Luis F, Rebeca C, Javier GP, Daniel A, Roberto H (2015) Adaptive stacked generalization for multiclass motor imagery-based brain computer interfaces. IEEE Transac Neural Syst Rehabil Eng 23(4): 702-712.

9. Satapathy Santosh Kumar, Loganathan D (2021) Prognosis of automated sleep staging based on two-layer ensemble learning stacking model using single-channel EEG signal. Soft Computing 25(24): 15445-15462.

10. Li Xiaowei, Bin H, Qunxi D, William C, Philip M, et al. (2011) EEG-based attention recognition. 2011 6th International Conference on Pervasive Computing and Applications. IEEE.

11. Mo H, Zhao Y (2016) Motor Imagery Electroencephalograph Classification Based on Optimized Support Vector Machine by Magnetic Bacteria Optimization Algorithm. Neural Process Lett 44(1): 185-197.

12. Sun S, Zhang C, Zhang D (2007) An experimental evaluation of ensemble methods for EEG signal classification. Pattern Recognit Lett 28(15): 2157-2163.

13. Adeli H, Zhou Z, Dadmehr N (2007) Analysis of EEG records in an epileptic patient using wavelet transform. J Neurosci. Methods 123(1): 69-87.

14. Lotte F (2014) A Tutorial on EEG Signal Processing Techniques for Mental State Recognition in Brain-Computer Interfaces. IEEE Trans Affect Comput 5(3): 327-339.

15. Ke X, Meng Q, Finley T, Wang T, Chen Y, et al. (2017) Light GBM: A Highly Efficient Gradient Boosting Decision Tree. Adv Neural Info Processing Systems.

16. Fan Junliang, Xia M, Lifeng W, Fucang Z, Xiang Y, et al. (2019) Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. Agric Water Manag 225: 105758.

17. Choubey, Hemant, and Alpana Pandey (2019) A new feature extraction and classification mechanisms For EEG signal processing. Multidimensional Systems and Signal Processing 30: 1793-1809.

18. Zhou Zhi-Hua (2012) Ensemble methods: foundations and algorithms. CRC press.

19. Singh Jaiteg, Farman A, Rupali G, Babar S, Daehan K (2023) A Survey of EEG and Machine Learning based methods for Neural Rehabilitation. IEEE Access.

20. Rasheed, Khansa, Adnan Q, Junaid Q, Shobi S, Patrick K, et al. (2020) Machine learning for predicting epileptic seizures using EEG signals: A review. IEEE Rev Biomed Engineer 14: 139-155.

Your next submission with Juniper Publishers will reach you the below assets
• Quality Editorial service
• Swift Peer Review
• Reprints availability
• E-prints Service
• Manuscript Podcast for convenient understanding
• Global attainment for your research
• Manuscript accessibility in different formats (Pdf, E-pub, Full Text, Audio)
• Unceasing customer service
Track the below URL for one-step submission
https://juniperpublishers.com/online-submission.php