Enhancing MI EEG Signal Classification With a Novel Weighted and Stacked Adaptive Integrated Ensemble Model: A Multi-Dataset Approach

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 June 2024, accepted 22 July 2024, date of publication 29 July 2024, date of current version 6 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3434654

RESEARCH ARTICLE

Enhancing MI EEG Signal Classification With a Novel Weighted and Stacked Adaptive Integrated Ensemble Model: A Multi-Dataset Approach

HOSSEIN AHMADI AND LUCA MESIN
Mathematical Biology and Physiology, Department of Electronics and Telecommunications, Politecnico di Torino, 10129 Turin, Italy
Corresponding author: Luca Mesin (luca.mesin@polito.it)

ABSTRACT Electroencephalography (EEG) based Brain-Computer Interfaces (BCIs) are vital for various applications, yet achieving accurate EEG signal classification, particularly for Motor Imagery (MI) tasks, remains a significant challenge. This study introduces a novel Weighted and Stacked Adaptive Integrated Ensemble Classifier (WS-AIEC), employing a comprehensive approach across six MI EEG datasets with 16 diverse Machine Learning (ML) classifiers. Through evaluations that encompass metric-based comparisons and learning curve analyses, we systematically ranked and clustered the classifiers. The WS-AIEC integrates the top-performing classifiers from each cluster and employs a unique blend of weighted and stacked ensemble techniques. Our results demonstrate the WS-AIEC's superior performance, achieving an exceptional accuracy of 99.58% on the BNCI2014-002 dataset and an average improvement of 20.23% in accuracy over the top-performing individual classifiers across all datasets. This significant enhancement underscores the innovative approach of our WS-AIEC in EEG signal classification for BCIs, setting a new benchmark for accuracy and reliability in the field.

INDEX TERMS Brain-computer interface, stacking ensemble models, weighted ensemble techniques, time series cross-validation, EEG signal processing, motor imagery EEG classification, ensemble learning.

I. INTRODUCTION

In the burgeoning field of interface technologies between the human brain and computers, advancements are transforming rehabilitation methods and the way we engage with digital environments [1]. At the heart of these innovations is the capability to decipher the intricate signals emanating from our brains, with Electroencephalography (EEG) taking center stage [2]. This technique is celebrated for its non-invasive approach, affordability, and adeptness at capturing the brain's dynamics in real-time [3]. A particularly fascinating application of EEG within this domain is the concept of Motor Imagery (MI) – vividly imagining a movement without physically performing it. This approach is critical for seamlessly integrating Brain-computer Interfaces (BCIs) across various fields, demanding precise interpretation of MI-related EEG signals [4]. The complexity of such data, coupled with potential environmental noise and the unique brainwave patterns of individuals, presents a significant challenge [5]. To navigate this, applying Machine Learning (ML) technologies has become prevalent, offering solutions that adapt to the specific nuances of each task and data set. However, their efficacy can be as variable as the data they seek to decode [6].

Despite the undeniable advantages of employing ML classifiers for the classification of MI EEG signals, the quest to identify the optimal classifier for a specific task and dataset emerges as a formidable challenge. The landscape of ML classifiers is vast, with each classifier boasting distinct characteristics and suitability for different types of EEG

The associate editor coordinating the review of this manuscript and approving it for publication was Muammar Muhammad Kabir.

data and tasks. This diversity, while beneficial, introduces complexity in selecting the most effective approach for signal interpretation [7]. The challenge is compounded by the intricate variability of brainwave patterns among individuals and potential environmental noise that can affect signal quality. Consequently, even when applying advanced optimization techniques to enhance classifier performance, pinpointing the optimal ML classifier for a given task remains time-consuming, resource-intensive, and uncertain. Therefore, achieving the highest possible accuracy in decoding MI EEG signals requires a meticulous and nuanced approach to classifier selection, acknowledging the likelihood that a universally optimal solution may be elusive given the current technological and methodological constraints.

While the field boasts numerous studies on MI EEG signal classification employing various ML classifiers, with a range of performances that sometimes approach near-perfection, a critical and overarching challenge remains largely unaddressed. The high-performance metrics reported in many studies are often the result of selecting classifiers, or a combination thereof, mainly by chance rather than through a systematic approach tailored to the specific characteristics of the data at hand. This lack of methodical selection is further compounded by a significant issue: the adaptability of these classifiers to new, unseen datasets. While notable, the impressive results achieved by some studies may not necessarily hold when applied to different datasets, leading to a dramatic decline in performance. This predicament underscores a fundamental flaw in the absence of a systematic, data-driven methodology for classifier selection. Such an approach would not only consider the inherent properties of the dataset but also its potential variability and the generalizability of the classification model to new data. As a result, the reliability of these high-performance outcomes is questioned, suggesting that they may be more a matter of fortunate coincidence than of robust, repeatable scientific discovery. The emphasis, therefore, shifts towards developing effective strategies that are adaptable and resilient across diverse data types, ensuring reliability and broader applicability in the realm of MI EEG signal classification.

One of the most effective solutions for addressing the challenges outlined above is using ensemble models. These models combine the capabilities of various classifiers to capitalize on each other's strengths, mitigating the limitations of individual models [8]. By doing so, ensembles can provide a nuanced analysis of EEG signals, drawing on the unique advantages of each constituent classifier, such as varied sensitivities to features and robustness against different types of noise. In practice, this synergistic approach means that where one classifier might falter, another in the ensemble will likely succeed, ensuring a more consistent and reliable overall performance [9]. Thus, ensemble models are a promising avenue for enhancing the accuracy of MI EEG signal analysis, marking a step forward in pursuing more sophisticated and versatile ML applications in the realm of EEG data interpretation [10].

In the ongoing effort to improve ML solutions in EEG analysis, numerous studies have explored the concept of ensemble learning, each adopting a distinct strategy to harness the collective power of classifiers. These explorations have repeatedly demonstrated the efficacy of ensembles in boosting accuracy. Nevertheless, they also reveal the need for a systematic approach to selecting classifiers that can adeptly navigate the vast variability of EEG datasets, ensuring robust performance across different conditions.

An ensemble of Deep Learning (DL) models, utilizing soft voting to integrate diverse architectures for optimal classification of MI EEG signals, showcased the power of combining multiple approaches [11]. In [12], the authors focused on ensemble models that improve MI EEG signal classification for brain cyborg applications by employing techniques like Boosting, Bagging, and Random Subspace, demonstrating their efficacy using time-frequency features. In another study [13], the authors combined stacking ensemble learning with Graph Convolutional Neural Networks (GCNNs), aiming to enhance MI task classification through a novel approach that integrates structural and functional connectivity. A subject-specific mental workload classifier using an ensemble of Convolutional Neural Networks (CNNs), each focusing on a different EEG channel, to better capture spatial information and enhance classification accuracy was proposed in [14]. An ensemble learning approach is discussed in [15], where majority voting decodes multi-class MI EEG signals by integrating Filter Bank Common Spatial Patterns (FBCSP) with a strategic classification extension. In [16], advanced classification methods in their ensemble model are integrated to improve EEG signal classification for MI tasks. In another study [17], the potential of an ensemble of classifiers for subject-independent binary classification of MI experiments demonstrated the ensemble's superior predictive performance. In [18], an ensemble learning method is introduced, combining Common Spatial Pattern (CSP) with ensemble strategies to mitigate the challenges posed by EEG signal nonstationarity and limited training data sizes. In [19], a novel framework combining ensemble learning with functional connectivity estimators demonstrates the advantages of integrating various classifiers to discern mental states.

Despite these innovative approaches, a gap remains in establishing a universally applicable, systematic methodology for selecting and combining classifiers within an ensemble to ensure robust performance across new and unseen datasets. While these referenced works represent significant strides in the field, their methodologies often reflect a trial-and-error approach rather than a scalable, data-driven framework adaptable to the vast variability of EEG data. This observation underscores the need for further research into developing a comprehensive strategy that capitalizes on the strengths of ensemble models and provides a reliable pathway for their application to diverse scenarios.

Our previous work [20] introduced the Correlation-Optimized Weighted Stacking Ensemble (COWSE) model.

The core innovation of the COWSE model lies in its sophisticated integration of 16 ML classifiers through a weighted stacking approach. This approach meticulously balances the strengths and weaknesses of each classifier based on an in-depth error correlation analysis and performance metrics evaluation across benchmark datasets. This process is complemented by employing a meta-classifier trained on the weighted predictions of the base classifiers. This novel methodology enabled the COWSE model to achieve an exceptional classification accuracy of 98.16%. The success of the COWSE model underscores the potential of integrating multiple ML classifiers to navigate the intricate patterns of EEG data, encouraging further exploration into advanced ensemble learning strategies.

Inspired by the achievements of our previous work and its exceptional results, we extend its innovative methodology to enhance further the classification of MI EEG signals. In this advancement, we incorporate the same 16 ML classifiers previously employed. To broaden the scope and generalize our findings, we integrate two additional datasets into our analysis, extending the evaluation beyond the original four datasets. This expansion aims to validate the applicability and robustness of our methodology across a more comprehensive array of data scenarios, ensuring its adaptability and relevance in various contexts.

In this novel development, we introduce a significant enhancement to the model's adaptability and effectiveness: the transition from static to dynamic weighting of classifiers within the ensemble. Unlike in our previous work, where the weights assigned to the base classifiers were fixed and unchanged during ensemble training, our new approach employs a dynamic weighting strategy. This method adapts the contribution of each classifier based on ongoing performance evaluations, allowing the ensemble to adjust and optimize its composition in response to the characteristics of the data being processed.

Our investigation illuminates a critical gap—the absence of a holistic integration of adaptive strategies with weighted and stacking techniques, further enriched by a systematic clustering process. This gap underscores the potential of a more sophisticated ensemble model capable of adeptly navigating EEG data's inherent complexities and variabilities. Remarkably, our study introduces the Weighted and Stacked Adaptive Integrated Ensemble Classifier (WS-AIEC), a model that intricately combines these methodologies. By leveraging the unique strengths of weighted approaches for individual classifier optimization, stacking methods for meta-level prediction refinement, and clustering techniques for intelligent classifier selection, the WS-AIEC model marks a significant advancement in the field. This integrative approach capitalizes on each method's synergistic strengths and addresses the critical need for adaptive ensemble models. Thus, the WS-AIEC model propels the field towards a new paradigm, establishing a comprehensive framework to enhance the accuracy and reliability of MI EEG-based BCIs.

In this study, our primary objective is to develop an ensemble model that integrates multiple classifiers to leverage their combined strengths. This approach differs from evaluating individual classifiers against the best models in the literature, as we aim to demonstrate that our ensemble model can achieve superior overall performance through strategic synergy. Thus, we focus on optimizing the ensemble rather than enhancing individual classifiers.

We focus on assessing the performance of various ML classifiers applied to MI EEG signals, explicitly focusing on conventional ML techniques. Unlike DL approaches, these traditional methods require an initial feature extraction phase before classification [21]. Therefore, we concentrated on employing ML models, deliberately omitting DL models due to the divergence in our research objectives.

While the surveyed studies may use a broad array of feature extraction techniques, our study employs CSP for feature extraction. However, a comparative analysis of feature extraction methodologies is not the central aim of our investigation. The "No Free Lunch" theorem [22], which postulates that no single approach consistently outperforms others in pattern classification, underscores the importance of understanding the task's nature and the dataset's specifications when selecting the most effective classifier.

This work introduces the WS-AIEC, a novel EEG signal classification approach within the BCIs domain, with our contributions being threefold:

- **First**, we present a comprehensive evaluation of 16 diverse ML classifiers across six MI EEG datasets, employing a rigorous metric-based comparison and learning curve analyses.
- **Second**, we develop a unique ensemble model that integrates the top-performing classifiers within each cluster through an innovative combination of weighted and stacked ensemble techniques, dynamically optimizing classifier contributions based on performance.
- **Lastly**, we demonstrate the superior performance of WS-AIEC, achieving unprecedented accuracy levels on multiple benchmark datasets, thus marking a significant advancement in EEG-based BCIs.

Following the Introduction, the Methodology section presents our experimental setup, detailing the datasets, classifiers, and our innovative ensemble model. The Results and Discussion sections provide an exhaustive analysis of the results, discussing their implications and potential limitations. The Conclusion section summarises our findings, highlights our proposed model's performance and limits, and identifies promising avenues for future research.

II. METHODOLOGY

This study's methodology is designed to evaluate and integrate a diverse array of ML classifiers systematically. Our approach encompasses a comprehensive evaluation of classifiers across multiple datasets, the development of a novel algorithm for optimal ensemble construction, and a rigorous assessment of the ensemble model's performance.

This section delineates the various components of our methodology, including the datasets and the classifiers used, the evaluation of the classifiers, and the design of the ensemble model.

A. DATASETS

The selection and analysis of datasets play a pivotal role in the robust evaluation of ML classifiers for MI EEG signal classification. A comprehensive and diverse set of datasets ensures that the classifiers and the proposed ensemble model are tested across various conditions, including varying subjects, channels, classes, trial durations, and sampling rates. Such diversity is crucial for assessing the generalizability and adaptability of the classification methods to different scenarios.

This study has meticulously chosen six datasets, each representing a unique configuration of parameters that challenge and evaluate the classifiers' capabilities. Each dataset has been selected based on its prevalence in MI EEG research, contribution to a comprehensive evaluation, and the opportunity it presents to advance our understanding of classifier performance in BCI technology. Table (1) provides an overview of the datasets utilized in this study, highlighting their specific characteristics and configurations.

B. CLASSIFIER SELECTION AND DESCRIPTION

Selecting individual classifiers is paramount in constructing a robust ensemble model. For this study, our classifier selection is informed by an in-depth analysis of each classifier's performance in handling high-dimensional neural data and their historical efficacy in MI EEG tasks. This strategic selection aims to capitalize on each classifier's unique advantages to the ensemble, from linear models adept at simplifying complex relationships through straightforward decision boundaries to non-linear classifiers capable of navigating the intricate structures within EEG signals. This phase of our study introduces a nuanced perspective on classifier integration, focusing on the individual merits of each model and their synergistic potential when combined within an ensemble framework. Our ensemble model seeks to leverage:

1) LINEAR MODELS

Efficiency and interpretability, crucial for understanding the underlying patterns within EEG signals. This category includes:

- Logistic Regression (LR)
- Linear Discriminant Analysis (LDA)
- Perceptron (PC)
- Stochastic Gradient Descent (SGD)
- Ridge Classifier (RC)
- Support Vector Machines (SVM)

2) NON-LINEAR AND INSTANCE-BASED LEARNING

Flexibility in capturing complex, non-linear interactions between features, comprising:

- k-Nearest Neighbors (KN)
- SVM with Radial Basis Function (SVM-rbf)

3) DECISION TREE-BASED MODELS

A structural approach to decision-making, offering insights into feature importance and data segmentation, including:

- Decision Trees (DT)
- Random Forest (RF)
- Extra Trees (ET)

4) BOOSTING MODELS

Sequential refinement of predictions, enhancing adaptability and learning from diverse data representations:

- Gradient Boosting (GB)
- AdaBoost (AB)

5) NEURAL NETWORKS

Advanced level of abstraction and feature extraction, represented by:

- Multi-layer Perceptron (MLP)

6) QUADRATIC MODELS

Handling more complex decision boundaries with models like:

- Quadratic Discriminant Analysis (QDA)

7) NAIVE BAYES CLASSIFIER

Utilizing probabilistic approaches, this category includes:

- Naive Bayes (NB)

We aim to underscore the ensemble's capacity to incorporate diverse learning strategies through this classifier categorization, enhancing its performance and generalizability.

C. PREPROCESSING, FEATURE EXTRACTION, AND DATA SPLITTING

The preprocessing and feature extraction phases are critical for preparing the MI EEG data for classification, ensuring that the input to the ML models is of high quality and relevance. This subsection outlines the steps to preprocess the data, extract meaningful features, and split the datasets for training/validation and testing purposes.

1) PREPROCESSING

The raw EEG signals were first subjected to a band-pass filtering process to enhance the signal quality by reducing noise and focusing on the frequency bands most relevant to MI tasks. Utilizing the MNE-Python toolbox [29], a finite impulse response (FIR) filter with a Hamming window was applied, selectively allowing frequencies in the 7-30 Hz range to pass. This frequency range was chosen to capture the alpha (8)-13 Hz) and beta (14-30 Hz) bands, which are significantly involved in MI activities. This selection is supported by extensive literature, including [30] and [31], which show that the 8-30 Hz range is commonly used in MI studies, reinforcing the relevance and appropriateness

TABLE 1. Datasets.

Dataset	Subjects	Channels	Classes	Trials / Class	Trial Duration	Sampling rate	Sessions
BNCI2014_001 [23]	9	22	4	144	4s	250Hz	2
BNCI2014_002 [24]	14	15	2	80	5s	512Hz	1
BNCI2014_004 [25]	9	3	2	360	4.5s	250Hz	5
BNCI2015_001 [26]	12	13	2	200	5s	512Hz	2
Zhou2016 [27]	4	14	3	160	5s	250Hz	3
AlexMI [28]	8	16	3	20	3s	512Hz	1

of our chosen frequency range. The filter parameters were carefully adjusted for each dataset to accommodate the EEG data's specific characteristics and recording conditions. The stop band frequencies were set to begin at 5 Hz and extend to 32 Hz, ensuring a clear delineation from the passband. The filter was designed to achieve a minimum attenuation of -40 dB in these stop bands, which helps in effectively suppressing frequencies outside the desired range. Additionally, the transition bands were defined from 5-7 Hz and 30-32 Hz, allowing for a smooth transition while maintaining the integrity of the passband frequencies.

2) FEATURE EXTRACTION

Following the preprocessing step, the CSP algorithm was employed for feature extraction. CSP is renowned for its efficacy in enhancing the signal-to-noise ratio by optimizing the variance differentiation between contrasting classes. This optimization amplifies the distinctiveness of the signals associated with various MI tasks and significantly improves the spatial pattern recognition critical for accurate classification [32].

3) DATA SPLITTING

To accurately evaluate the performance of our ensemble model and individual classifiers, we followed a structured approach to data splitting that respects the temporal dependencies inherent in EEG data. For each dataset, all trials from all classes and sessions were first combined and then sorted chronologically within each subject. The chronologically sorted data for each subject was then split into 80% for training/validation and 20% for testing. Within the 80% training/validation split, we employed Time Series Cross-Validation (TSCV) using scikit-learn to ensure that the temporal order of the data was maintained. This method prevents future data from being used in the training process, maintaining the integrity of the temporal sequence. The results presented in this study are the average accuracies calculated across all subjects, ensuring that the reported performance metrics reflect a comprehensive evaluation of the ensemble model's effectiveness for each individual subject.

D. MODEL TRAINING, VALIDATION, AND EVALUATION

Our model training, validation, and evaluation approach is designed to optimize classifier performance and rigorously test the ensemble model constructed from these classifiers. Initially, as detailed in Data Splitting, the datasets were partitioned into two subsets for a comprehensive model assessment and refinement environment.

A pivotal component of our methodology involved fine-tuning the cross-validation strategy to fit our data's temporal nature perfectly. Through an extensive grid search hyper-parameter optimization process, which included evaluating the number of folds in TSCV, we identified the optimal configuration as an 8-fold TSCV. This setup ensured that each data segment was utilized efficiently for training and validation across different phases, providing a holistic view of each classifier's performance under varying conditions.

Furthermore, our study delved into the influence of training data volume on the effectiveness of the classifiers. By incrementally increasing the amount of data used for training from 10% to the entirety of the training/validation set, we could discern the relationship between data volume and classification accuracy.

Following the training phase, our evaluation involved an in-depth analysis of the classifiers' performance on the dedicated testing set. Learning curves were meticulously plotted to visualize and understand the learning progression across training iterations. These curves, alongside other performance metrics, played a crucial role in the comparative analysis of the classifiers, facilitating a data-driven selection process for the ensemble model.

The innovation of our methodology shines in the ensemble model's construction. Instead of a straightforward selection of top-performing classifiers, our approach employed a nuanced clustering technique. Classifiers were grouped based on their performance metrics across various datasets, employing hierarchical clustering to identify patterns of similarity. This strategy allowed us to cherry-pick the most efficient classifier from each cluster, assembling an ensemble that embodies a harmonious balance of diversity and peak performance.

As the culmination of our methodology, the ensemble model was benchmarked against individual classifiers to evaluate its superiority. This comparative analysis was essential not only to showcase the ensemble's efficacy but also to validate the effectiveness of our novel ensemble construction strategy.

E. MULTIFACETED CLASSIFIER EVALUATION FOR ENSEMBLE MODEL OPTIMIZATION

We implemented two distinct approaches for ranking the classifiers to construct an effective ensemble model, ensuring a comprehensive evaluation based on varied criteria. These approaches are detailed below.

1) COMPREHENSIVE PERFORMANCE ASSESSMENT AND RANKING OF THE CLASSIFIERS

In the first approach, we focused on the actual performance metrics of the classifiers across the six datasets used in our study. We calculated average scores for accuracy, precision, recall, F1 score, Area Under the Receiver Operating Characteristic Curve (AUC-ROC), and kappa. The classifiers were then ranked based on an overall score derived from these metrics, directly comparing their effectiveness in our context. We calculated each classifier's overall performance score S_i based on multiple metrics. Let $w_{A_i}, w_{Pr_i}, w_{R_i}, w_{F1_i}, w_{AUC_i}, w_{K_i}$ be the weights for accuracy, precision, recall, F1 score, AUC-ROC, and kappa, respectively. The overall performance score S_i for each classifier is then:

$$S_i = w_{A_i} \cdot A_i + w_{Pr_i} \cdot Pr_i + w_{R_i} \cdot R_i + w_{F1_i} \cdot F1_i + w_{AUC_i} \cdot AUC_i + w_{K_i} \cdot K_i \quad (1)$$

where

- A_i is the accuracy of classifier i;
- Pr_i is the precision of classifier i;
- R_i is the recall of classifier i;
- $F1_i$ is the F1 score of classifier i;
- AUC_i is the AUC-ROC of classifier i;
- K_i is the kappa score of classifier i.

In the current analysis, each performance metric is given equal importance, reflecting a balanced approach to classifier assessment. However, the methodology is designed to be adaptable, allowing for assigning different weights to metrics as dictated by their significance to specific tasks within the domain.

2) EVALUATION OF CLASSIFIER EFFICIENCY THROUGH LEARNING CURVE ANALYSIS

The second approach in our study evaluates the classifiers' learning dynamics by examining the learning curves. These curves reveal how well a model generalizes from the training data to unseen data as the number of training examples increases. Three key parameters derived from the learning curves—Area Under the Curve of Cross-Validation (AUC-CV) scores, Convergence Rates (CR), and Performance Stability (PS)—were analyzed to rank classifiers based on their learning efficiency and stability over time.

- **AUC-CV:** The AUC-CV represents the classifier's ability to maintain high performance across different training data sizes. It is calculated as the area under the learning curve plot, with the x-axis representing the number of training examples and the y-axis representing the cross-validation score. A higher AUC-CV indicates a model that learns effectively and performs well across varying training data. Mathematically, it is given by the integral:

$$AUC_{CV,i} = \int_a^b CV_i(t)dt \quad (2)$$

where a and b represent the minimum and maximum training sizes, respectively, and $CV_i(t)$ is the continuous function of cross-validation scores of classifier i across training size t.

- **CR:** CR measures the rate at which the classifier's training score converges to its validation score, indicative of the model's ability to generalize from the training to the validation set. A lower CR suggests that the classifier is learning generalizable patterns more quickly. It is defined as the absolute difference between the final training score mean, and the final cross-validation score mean:

$$CR_i = |\text{Mean}(TSM_i) - \text{Mean}(CVM_i)| \quad (3)$$

where TSM_i is training score mean and CVM_i is cross-validation score mean of classifier i at the largest training size considered.

- **PS:** PS quantifies the variability in the classifier's performance across different training sizes, with a lower PS indicating a more stable performance. It is the average standard deviation of the cross-validation scores:

$$PS_i = \frac{1}{n}\sum_{j=1}^{n} \sigma_{CV_{i,j}} \quad (4)$$

where $\sigma_{CV_{i,j}}$ is the standard deviation of the cross-validation scores at the j-th training size, and n is the total number of training sizes analysed.

Weights w_{AUC}, w_{CR}, w_{PS} were assigned to each parameter to reflect their relative importance, given the specific goals of the study and the characteristics of the task at hand. In this analysis, equal weights were used to avoid biasing the model towards any particular aspect of learning efficiency or stability:

$$L_i = w_{AUC} \cdot AUC_{CV,i} + w_{CR} \cdot CR_i + w_{PS} \cdot PS_i \quad (5)$$

This combined score for learning curve parameters L_i is a comprehensive measure of a classifier's effectiveness, considering its performance, learning rate, and stability, which are crucial for real-world applications where data variability and volume can be unpredictable.

F. CLASSIFIER CLUSTERING AND OPTIMAL SELECTION FOR ENSEMBLE CONSTRUCTION

To construct our ensemble model strategically, we employed clustering techniques to group classifiers based on their performance characteristics across multiple evaluative dimensions. Hierarchical clustering, guided by the Euclidean distance metric, allowed us to identify natural groupings among classifiers:

$$D(i,j) = \sqrt{\sum_{k=1}^{m}(f_{i,k} - f_{j,k})^2} \quad (6)$$

where $D(i,j)$ quantifies the similarity between classifiers i and j, with $f_{i,k}$ and $f_{j,k}$ representing their performance

feature values and m being the number of considered features. By analyzing distances in this multidimensional feature space, we can discern clusters of classifiers that perform similarly, facilitating a more nuanced selection for our ensemble.

The optimal number of clusters was determined using the Elbow method, which examines the within-cluster sum of squares versus the number of clusters. This analysis is crucial for balancing diversity and redundancy in the ensemble's decision-making process.

Within each identified cluster, we selected the best-performing classifier based on an average rank that integrates performance metrics ranking and learning curves parameters ranking:

$$\bar{R}_i = w_S \cdot R_{Si} + w_{LC} \cdot R_{LCi} \tag{7}$$

where

- R_{Si} is the rank of classifier i based on direct performance metrics;
- R_{LCi} is the rank of classifier i based on learning curves parameters ranking.

Weights w_S and w_{LC} correspond to the importance of direct performance metrics and learning curve parameters, respectively. The chosen classifiers are then weighted inversely proportional to their average rank $\bar{R}_i$:

$$w_i = \frac{1}{\bar{R}_i} \tag{8}$$

This static weight calculation is pivotal for the initial phase of the ensemble model construction, specifically in strategically selecting the meta-classifier and the base classifiers. The weight w_i assigned to each classifier serves a dual purpose:

- **Meta-Classifier Selection:** The classifier with the highest static weight (i.e., the lowest average rank) is designated as the meta-classifier. This classifier is deemed to have the most robust performance across multiple datasets and metrics, making it ideal for integrating the predictions from the base classifiers.
- **Base Classifier Selection:** The remaining classifiers, ranked according to their static weights, are selected as base classifiers. While diverse in their methodological approaches and performance characteristics, these classifiers contribute valuable predictive perspectives to the ensemble.

It's imperative to note that this static weight assignment is solely for the initial selection and categorization of the meta and base classifiers. For the base classifiers, post-selection, we transition to a dynamic weight adjustment mechanism during the ensemble's training phase. This approach allows the ensemble to adaptively recalibrate the influence of each base classifier based on real-time performance metrics, specifically accuracy on a validation subset. The dynamic weighting formula introduced in the subsequent section enables this live adjustment, ensuring the ensemble's predictions are continually optimized in response to the evolving performance landscape of the base classifiers. We establish a comprehensive and adaptive ensemble model by clearly distinguishing between using static weights for the initial selection of meta and base classifiers and applying dynamic weights for ongoing optimization of the base classifiers. This model leverages the individual strengths of diverse classifiers and maintains flexibility and responsiveness to data dynamics and classifier performance changes over time.

After evaluating individual classifiers, we introduce the WS-AIEC. This ensemble model is an innovative amalgamation of the strengths of both weighted and stacked approaches, designed to adaptively integrate classifier predictions in a framework suited for the intricacies of MI EEG data. As depicted in Figure 1, the methodology progresses through four distinct phases, starting with data preprocessing and culminating in the final comparative evaluation.

G. DYNAMIC WEIGHT ADJUSTMENT FOR BASE CLASSIFIERS

Building upon the framework outlined for classifier clustering and optimal selection, our ensemble model introduces a novel approach to dynamic weight adjustment for the base classifiers. This method aims to optimize the ensemble's performance by adjusting the contributions of the base classifiers in real time based on their predictive accuracy on a validation subset of the training data.

1) DYNAMIC WEIGHT CALCULATION

The dynamic weight adjustment process is executed during the training phase of the ensemble model. After selecting the base classifiers and the meta-classifier using the methodology described in the previous section, we further refine the ensemble by dynamically adjusting the weights of the base classifiers. This is achieved by monitoring the performance of each base classifier on a validation subset, which constitutes 20% of the training data reserved specifically for this purpose. The dynamic weight $w_{i,t}$ for classifier i at training epoch t is calculated as follows:

$$w_{i,t} = \frac{\exp(\alpha \cdot ACC_{i,t})}{\sum_{j=1}^{n} \exp(\alpha \cdot ACC_{j,t})} \tag{9}$$

where

- $ACC_{i,t}$ is the accuracy of classifier i on the validation subset at epoch t;
- α is a scaling parameter that controls the rate at which the weights adjust to the classifier's accuracy changes;
- n is the ensemble's total number of base classifiers.

This equation ensures that classifiers with higher accuracy on the validation subset receive a larger share of the overall vote in the ensemble prediction. The scaling parameter α allows for the adjustment of the sensitivity of the weights to changes in classifier performance, enabling the ensemble to adapt more quickly or more smoothly to the evolving data characteristics.

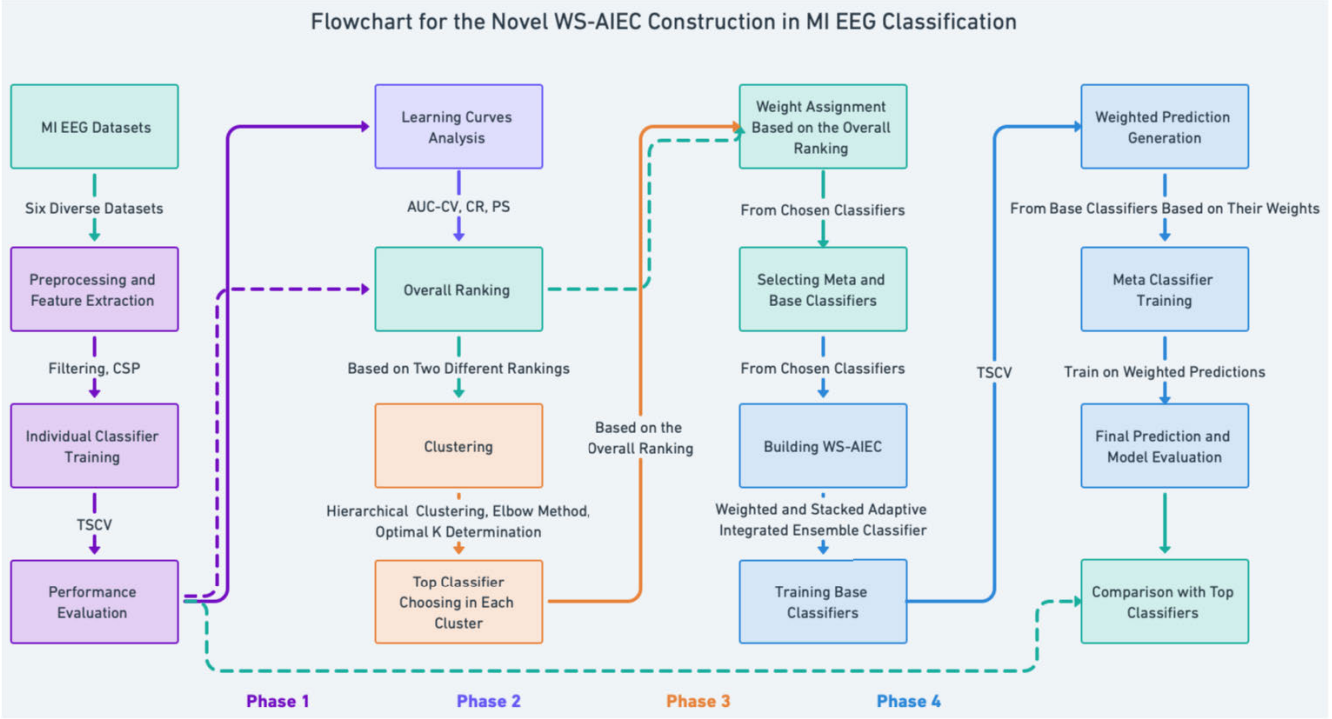

FIGURE 1. Sequential overview of the WS-AIEC methodology. The flowchart delineates a comprehensive four-phase process from initial data preparation to the final comparative evaluation. It encapsulates steps including preprocessing, classifier evaluation, ensemble construction using the novel WS-AIEC approach, and performance assessment against top benchmark classifiers.

To optimize the selection of α based on the unique characteristics of each dataset, we employ automated α tuning using Bayesian Optimization. This advanced method systematically explores the parameter space of α, evaluating its impact on ensemble performance through a guided search that balances the exploration of new parameter values with the exploitation of known high-performing values. By leveraging Bayesian Optimization, we ensure that α is tuned to the optimal setting that maximizes the predictive performance of the ensemble model while considering the specificities and variability inherent to each dataset.

This approach enhances the ensemble model's responsiveness to changes in base classifier performance and customizes the dynamic weight adjustment process to the unique features and challenges presented by the data, further optimizing the ensemble's accuracy and adaptability.

2) FEATURE MATRIX CONSTRUCTION

The dynamically weighted predictions of the base classifiers are aggregated to construct a feature matrix. This matrix serves as the input for the meta-classifier, effectively transforming the ensemble's diverse predictive signals into a cohesive dataset for final decision-making.

The feature matrix, denoted as F, is constructed by compiling the weighted predictions from each base classifier for all instances in the training dataset. Each row in F corresponds to a single instance, while each column represents the weighted prediction from one of the base classifiers. The formation of F is described by

$$F_{i,j} = w_{j,t} \cdot P_{i,j} \tag{10}$$

where

- $F_{i,j}$ is the element of the feature matrix corresponding to the i-th instance and the prediction from the j-th base classifier;
- $w_{j,t}$ is the dynamic weight of the j-th classifier at training epoch t, as recalculated at each epoch based on Equation (9);
- $P_{i,j}$ is the prediction of the j-th classifier for the i-th instance.

This matrix effectively captures the varying contributions of each base classifier to the ensemble's predictive power, dynamically adjusting to each classifier's evolving accuracy over the training process.

3) FINAL PREDICTION BY META-CLASSIFIER

The meta-classifier, having been trained on the feature matrix F, is equipped to make the final prediction by effectively integrating the insights provided by the base classifiers. The final ensemble prediction for a new instance is given by

$$\hat{y}_i = \text{Meta-Classifier}(F_i) \tag{11}$$

where

- $\hat{y}_i$ is the predicted label for the i-th instance by the ensemble;

- F_i is the i-th row of the feature matrix, representing the weighted predictions of all base classifiers for the i-th instance;
- Meta-Classifier denotes the prediction function of the meta-classifier.

The meta-classifier's ability to act on the dynamically generated feature matrix enables the ensemble model to leverage the collective strengths of its components. By adjusting the influence of each base classifier based on real-time performance and synthesizing their predictions, the ensemble achieves a high degree of adaptability and robustness. This method ensures that the ensemble model captures the diverse predictive signals from its base classifiers and remains responsive to data characteristics and classifier performance changes, leading to improved predictive accuracy on unseen data.

The algorithmic description presented in Algorithm 1 outlines the step-by-step process of constructing the WS-AIEC, from the initial data preparation to the final decision-making phase, emphasizing its adaptive weighting and stacking mechanism.

The effectiveness of the WS-AIEC will be rigorously evaluated and compared with top-performing individual classifiers from the initial ranking phase. The results section will detail this comparative analysis, highlighting the advantages of our adaptive ensemble approach.

Algorithm 1 WS-AIEC Construction With Dynamic Weight Assignments

1: Begin with the selection of diverse MI EEG datasets.
2: Preprocess data using band-pass filtering and CSP for feature extraction.
3: Split data into training/validation (80%) and testing sets (20%) using TSCV.
4: **for** each dataset **do**
5: Train classifiers using TSCV and plot learning curves.
6: Evaluate classifiers based on accuracy, precision, recall, F1 score, AUC-ROC, and kappa.
7: Evaluate classifiers based on AUC-CV, CR, and PS.
8: **end for**
9: Rank classifiers based on comprehensive performance assessment and learning curve analysis.
10: Apply hierarchical clustering to group classifiers based on two ranking approaches.
11: Determine the optimal number of clusters using the elbow method.
12: **for** each cluster **do**
13: Select the top-performing classifier in each cluster based on average rank from all evaluation approaches.
14: **end for**
15: Assign static weights inversely proportional to the overall ranking for initial classifier selection. The classifier with the highest rank is designated as the meta-classifier. The rest are selected as base classifiers.
16: Dynamically adjust weights of base classifiers using Bayesian Optimization based on their performance on the validation set during training.
17: Generate a feature matrix from the dynamically weighted predictions of the base classifiers.
18: Train the meta-classifier on this feature matrix.
19: Synthesize the final decision output by the meta-classifier.
20: Evaluate and compare the WS-AIEC performance with top-performing classifiers from the initial phase.
21: **return** The final decision and performance comparison results.

III. RESULTS

We commence the results with a direct comparison of classifier performances. This comparison is delineated in a series of subtables, each corresponding to a distinct dataset. The subtables summarise each classifier's performance metrics, including accuracy, precision, recall, F1-score, AUC-ROC, and kappa scores, as presented in Table 2.

Following the tabulated comparisons, we visually depict these performance metrics to illustrate classifiers' differences better. Figure 2 presents a comprehensive visual analysis of each classifier's accuracy across the datasets. These bar charts highlight the individual and average accuracies, offering a clear visual representation that complements the detailed data in the tables. This juxtaposition of visual and numerical data underlines the relative performance of classifiers, shedding light on those that consistently exceed average accuracy and those that do not meet the benchmark. Such insights are invaluable for informing subsequent decisions regarding classifier selection for the development of robust ensemble methods.

Proceeding with individual dataset analyses, we synthesize the results into a comprehensive overview, encapsulating the performance metrics across all datasets. This unified perspective is crucial for identifying broader trends and setting the groundwork for the ranking of classifiers. Table 3 provides this summary, offering a panoramic view of the classifiers' performance and facilitating their subsequent evaluation.

Advancing from the groundwork laid by the performance metrics, our focus shifts to the classifiers' learning curves, a vital component in assessing their efficiency. Learning curves yield insights into how classifier performance evolves by including increasingly more significant data subsets. They are crucial to gauging the trade-off between learning and overfitting and the classifiers' generalization ability.

For brevity and clarity, we present the learning curves from just one representative dataset, BNCI2014-002 (Figure 3), as an exemplar. While learning curves were generated for all datasets, they exhibit similar patterns and trends. The selected figure for BNCI2014-002 provides a detailed visual representation of how each classifier's performance

TABLE 2. Comparison of the classifiers based on their performances.

(a) BNCI2014-001

Classifier	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Kappa	Rank
MLP	0.6734	0.6793	0.6734	0.6729	0.7816	0.5640	1
SVM-rbf	0.6638	0.6709	0.6638	0.6623	0.7752	0.5516	2
LR	0.6619	0.6694	0.6619	0.6619	0.7739	0.5489	3
LDA	0.6600	0.6669	0.6600	0.6595	0.7724	0.5462	4
QDA	0.6561	0.6590	0.6561	0.6548	0.7696	0.5409	5
SVM	0.6542	0.6618	0.6542	0.6536	0.7687	0.5389	6
RC	0.6542	0.6560	0.6542	0.6506	0.7679	0.5378	7
RF	0.6351	0.6396	0.6351	0.6351	0.7552	0.5129	8
ET	0.6341	0.6403	0.6341	0.6338	0.7550	0.5119	9
GB	0.6312	0.6379	0.6312	0.6315	0.7529	0.5081	10
SGD	0.6255	0.6476	0.6255	0.6097	0.7509	0.5007	11
KN	0.6188	0.6354	0.6188	0.6194	0.7456	0.4922	12
NB	0.6073	0.6105	0.6073	0.6025	0.7360	0.4755	13
PC	0.5852	0.6452	0.5852	0.5556	0.7219	0.4448	14
DT	0.5527	0.5554	0.5527	0.5510	0.7002	0.4026	15
AB	0.5508	0.5650	0.5508	0.5460	0.6986	0.3994	16

(b) BNCI2014-002

Classifier	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Kappa	Rank
LR	0.7768	0.7852	0.7768	0.7768	0.7796	0.5515	1
SVM-rbf	0.7723	0.7861	0.7723	0.7719	0.7765	0.5449	2
RC	0.7723	0.7846	0.7723	0.7724	0.7758	0.5440	3
SVM	0.7723	0.7843	0.7723	0.7720	0.7754	0.5436	4
ET	0.7701	0.7823	0.7701	0.7699	0.7732	0.5388	5
MLP	0.7701	0.7805	0.7701	0.7700	0.7724	0.5379	6
LDA	0.7634	0.7778	0.7634	0.7631	0.7670	0.5278	7
SGD	0.7634	0.7735	0.7634	0.7564	0.7640	0.5203	8
GB	0.7589	0.7746	0.7589	0.7586	0.7639	0.5197	9
RF	0.7589	0.7703	0.7589	0.7588	0.7628	0.5168	10
DT	0.7589	0.7644	0.7589	0.7580	0.7585	0.5118	11
KN	0.7545	0.7708	0.7545	0.7546	0.7608	0.5117	12
AB	0.7500	0.7643	0.7500	0.7501	0.7545	0.4989	13
QDA	0.7478	0.7572	0.7478	0.7482	0.7505	0.4930	14
PC	0.7210	0.7543	0.7210	0.6900	0.7285	0.4505	15
NB	0.7143	0.7200	0.7143	0.7141	0.7153	0.4257	16

(c) BNCI2014-004

Classifier	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Kappa	Rank
SVM-rbf	0.7331	0.7400	0.7331	0.7312	0.7345	0.4677	1
LDA	0.7322	0.7387	0.7322	0.7315	0.7344	0.4664	2
MLP	0.7311	0.7359	0.7311	0.7304	0.7325	0.4631	3
LR	0.7297	0.7375	0.7297	0.7284	0.7329	0.4631	4
SVM	0.7304	0.7364	0.7304	0.7291	0.7322	0.4624	5
QDA	0.7304	0.7371	0.7304	0.7289	0.7320	0.4618	6
RC	0.7290	0.7360	0.7290	0.7282	0.7314	0.4605	7
RF	0.7296	0.7340	0.7296	0.7296	0.7305	0.4593	8
GB	0.7257	0.7304	0.7257	0.7257	0.7270	0.4523	9
NB	0.7173	0.7226	0.7173	0.7161	0.7183	0.4348	10
ET	0.7120	0.7151	0.7120	0.7122	0.7121	0.4228	11
AB	0.7079	0.7140	0.7079	0.7071	0.7104	0.4186	12
KN	0.7081	0.7116	0.7081	0.7082	0.7081	0.4152	13
DT	0.6885	0.6918	0.6885	0.6882	0.6876	0.3749	14
PC	0.6964	0.6855	0.6964	0.6498	0.6911	0.3830	15
SGD	0.6736	0.6955	0.6736	0.6300	0.6820	0.3634	16

(d) BNCI2015-001

Classifier	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Kappa	Rank
SVM-rbf	0.8312	0.8361	0.8312	0.8309	0.8325	0.6627	1
MLP	0.8299	0.8328	0.8299	0.8297	0.8304	0.6594	2
SVM	0.8274	0.8319	0.8274	0.8271	0.8287	0.6551	3
LR	0.8267	0.8301	0.8267	0.8266	0.8277	0.6534	4
RF	0.8205	0.8249	0.8205	0.8201	0.8213	0.6408	5
KN	0.8201	0.8248	0.8201	0.8197	0.8215	0.6407	6
RC	0.8191	0.8216	0.8191	0.8190	0.8195	0.6379	7
ET	0.8170	0.8233	0.8170	0.8163	0.8188	0.6344	8
LDA	0.8160	0.8191	0.8160	0.8158	0.8164	0.6318	9
GB	0.8153	0.8184	0.8153	0.8152	0.8164	0.6307	10
AB	0.8090	0.8121	0.8090	0.8090	0.8096	0.6180	11
QDA	0.7948	0.8007	0.7948	0.7928	0.7957	0.5899	12
NB	0.7917	0.8081	0.7917	0.7839	0.7920	0.5826	13
DT	0.7802	0.7828	0.7802	0.7800	0.7810	0.5603	14
SGD	0.7660	0.8010	0.7660	0.7557	0.7696	0.5363	15
PC	0.7535	0.8181	0.7535	0.7228	0.7578	0.5155	16

(e) Zhou2016

Classifier	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Kappa	Rank
ET	0.8019	0.8084	0.8019	0.8015	0.8517	0.7026	1
MLP	0.7910	0.7943	0.7910	0.7902	0.8442	0.6865	2
RF	0.7903	0.7959	0.7903	0.7895	0.8428	0.6850	3
SVM-rbf	0.7860	0.7946	0.7860	0.7849	0.8408	0.6792	4
KN	0.7796	0.7958	0.7796	0.7796	0.8365	0.6699	5
GB	0.7801	0.7883	0.7801	0.7807	0.8355	0.6696	6
SVM	0.7795	0.7856	0.7795	0.7788	0.8361	0.6695	7
LDA	0.7754	0.7792	0.7754	0.7745	0.8327	0.6628	8
LR	0.7745	0.7788	0.7745	0.7737	0.8319	0.6617	9
RC	0.7672	0.7698	0.7672	0.7654	0.8269	0.6509	10
PC	0.7410	0.8079	0.7410	0.7311	0.8120	0.6162	11
SGD	0.7417	0.7759	0.7417	0.7360	0.8111	0.6156	12
QDA	0.7381	0.7455	0.7381	0.7374	0.8035	0.6061	13
AB	0.7274	0.7352	0.7274	0.7252	0.7955	0.5902	14
DT	0.7048	0.7098	0.7048	0.7023	0.7783	0.5565	15
NB	0.6805	0.6882	0.6805	0.6745	0.7589	0.5186	16

(f) AlexMI

Classifier	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Kappa	Rank
KN	0.8333	0.8713	0.8333	0.8269	0.8661	0.7427	1
SVM	0.8056	0.8727	0.8056	0.8182	0.8489	0.7040	2
RC	0.8056	0.8389	0.8056	0.8025	0.8494	0.6998	3
SVM-rbf	0.7778	0.8505	0.7778	0.7837	0.8327	0.6649	4
MLP	0.7772	0.8227	0.7772	0.7784	0.8327	0.6596	5
LR	0.7770	0.8217	0.7770	0.7764	0.8321	0.6575	6
NB	0.7500	0.7847	0.7500	0.7601	0.7938	0.6111	7
ET	0.7222	0.8019	0.7222	0.7352	0.7994	0.5819	8
SGD	0.7222	0.7889	0.7222	0.7313	0.7929	0.5853	9
PC	0.7222	0.7403	0.7222	0.7222	0.7892	0.5689	10
RF	0.6944	0.7875	0.6944	0.7094	0.7827	0.5465	11
GB	0.6944	0.7431	0.6944	0.7009	0.7763	0.5376	12
LDA	0.6667	0.7569	0.6667	0.6887	0.7596	0.5000	13
DT	0.6667	0.7148	0.6667	0.6741	0.7434	0.4970	14
AB	0.6111	0.6501	0.6111	0.5775	0.5814	0.4429	15
QDA	0.5556	0.6954	0.5556	0.5529	0.7161	0.3808	16

evolves with increasing training data, encapsulating the trends generally representative across other datasets.

Table 4 quantifies the learning curve data for all datasets to complement this visual analysis and provide a comprehensive overview. Each subtable ranks the classifiers based on the critical metrics derived from their learning curves, including AUC-CV, CR, and PS, evaluated in raw and normalized forms for a thorough comparative analysis. The normalization process was applied to each metric to ensure a fair comparison across classifiers, and the ranks were calculated based on the average of these normalized scores.

Furthermore, for a consolidated perspective, Table 5 amalgamates the individual learning curve metrics into a ranking across all datasets. This approach offers a comparative view

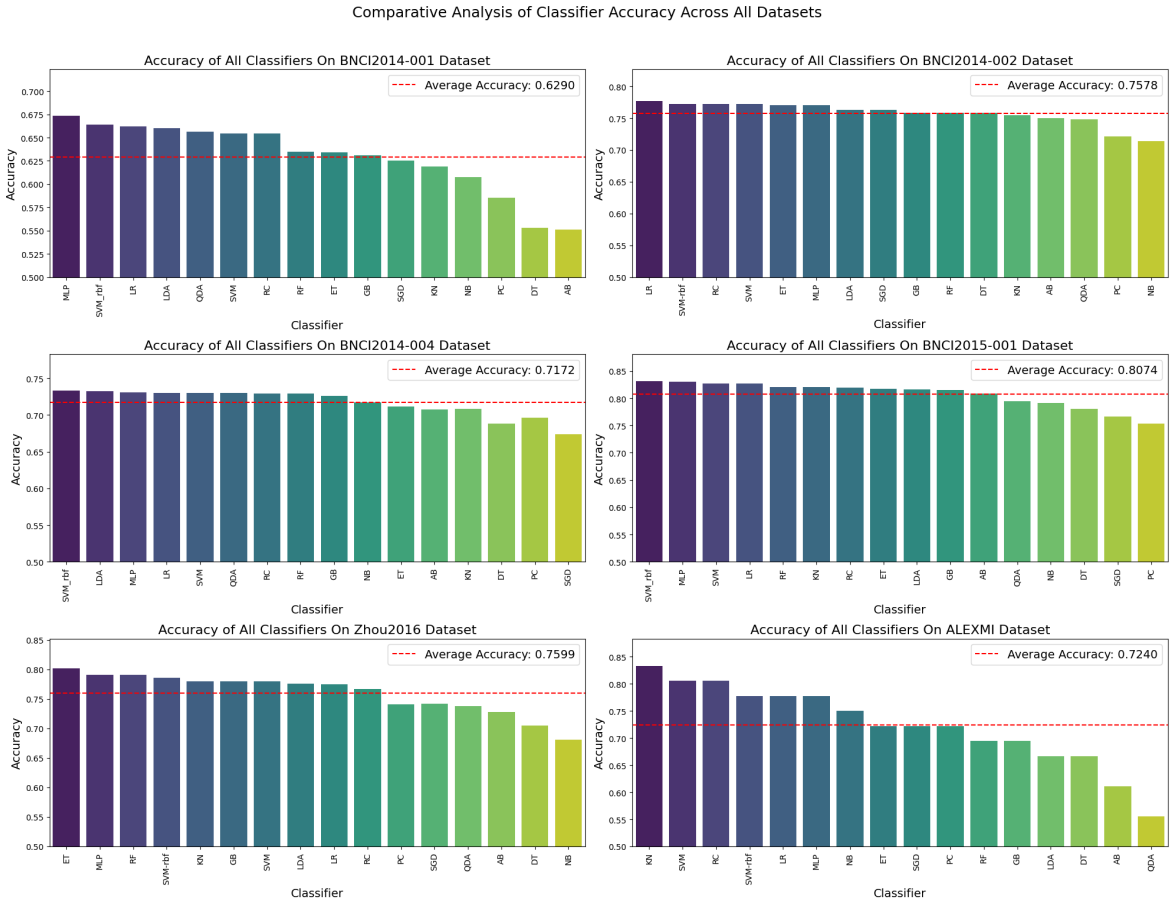

FIGURE 2. Comparative analysis of the accuracy for all classifiers across all datasets. Each chart plots the accuracy of individual classifiers, with the dashed line indicating the average accuracy across all classifiers for the respective dataset. These visualizations facilitate the identification of classifiers that consistently outperform the average and those that fall short, thereby informing decisions about classifier selection for ensemble methods.

TABLE 3. Aggregated performance metrics and ranking of classifiers across all datasets.

Classifier	Accuracy	Precision	Recall	F1 Score	AUC - ROC	Kappa	Overall Score	Rank
SVM	0.7616	0.7788	0.7616	0.7631	0.7983	0.5956	0.7432	1
SVM-rbf	0.7607	0.7797	0.7607	0.7608	0.7987	0.5952	0.7426	2
MLP	0.7622	0.7743	0.7622	0.7619	0.7990	0.5951	0.7424	3
LR	0.7579	0.7706	0.7579	0.7576	0.7965	0.5897	0.7384	4
RC	0.7579	0.7678	0.7579	0.7564	0.7952	0.5885	0.7373	5
KN	0.7524	0.7683	0.7524	0.7514	0.7898	0.5787	0.7322	6
ET	0.7429	0.7619	0.7429	0.7448	0.7850	0.5654	0.7238	7
RF	0.7381	0.7587	0.7381	0.7404	0.7826	0.5602	0.7197	8
LDA	0.7356	0.7564	0.7356	0.7389	0.7804	0.5558	0.7171	9
GB	0.7343	0.7488	0.7343	0.7354	0.7787	0.5530	0.7141	10
SGD	0.7154	0.7471	0.7154	0.7032	0.7618	0.5203	0.6938	11
QDA	0.7038	0.7325	0.7038	0.7025	0.7612	0.5121	0.6860	12
NB	0.7102	0.7224	0.7102	0.7085	0.7524	0.5081	0.6853	13
PC	0.7032	0.7419	0.7032	0.6786	0.7501	0.4965	0.6789	14
DT	0.6920	0.7032	0.6920	0.6923	0.7415	0.4839	0.6675	15
AB	0.6927	0.7068	0.6927	0.6858	0.7250	0.4947	0.6663	16

of the classifiers' ability to generalize from the data distilled into a singular, cohesive ranking system.

After evaluating our classifiers' individual and collective learning capabilities through detailed curve analysis, we now consolidate our findings. Table 6 provides an overarching view of the classifiers, amalgamating the insights from the various evaluation metrics into a single composite score.

Having established a comprehensive ranking system with Table 6, we now focus on the relational dynamics between classifiers through cluster analysis. Figure 4 provides a visual

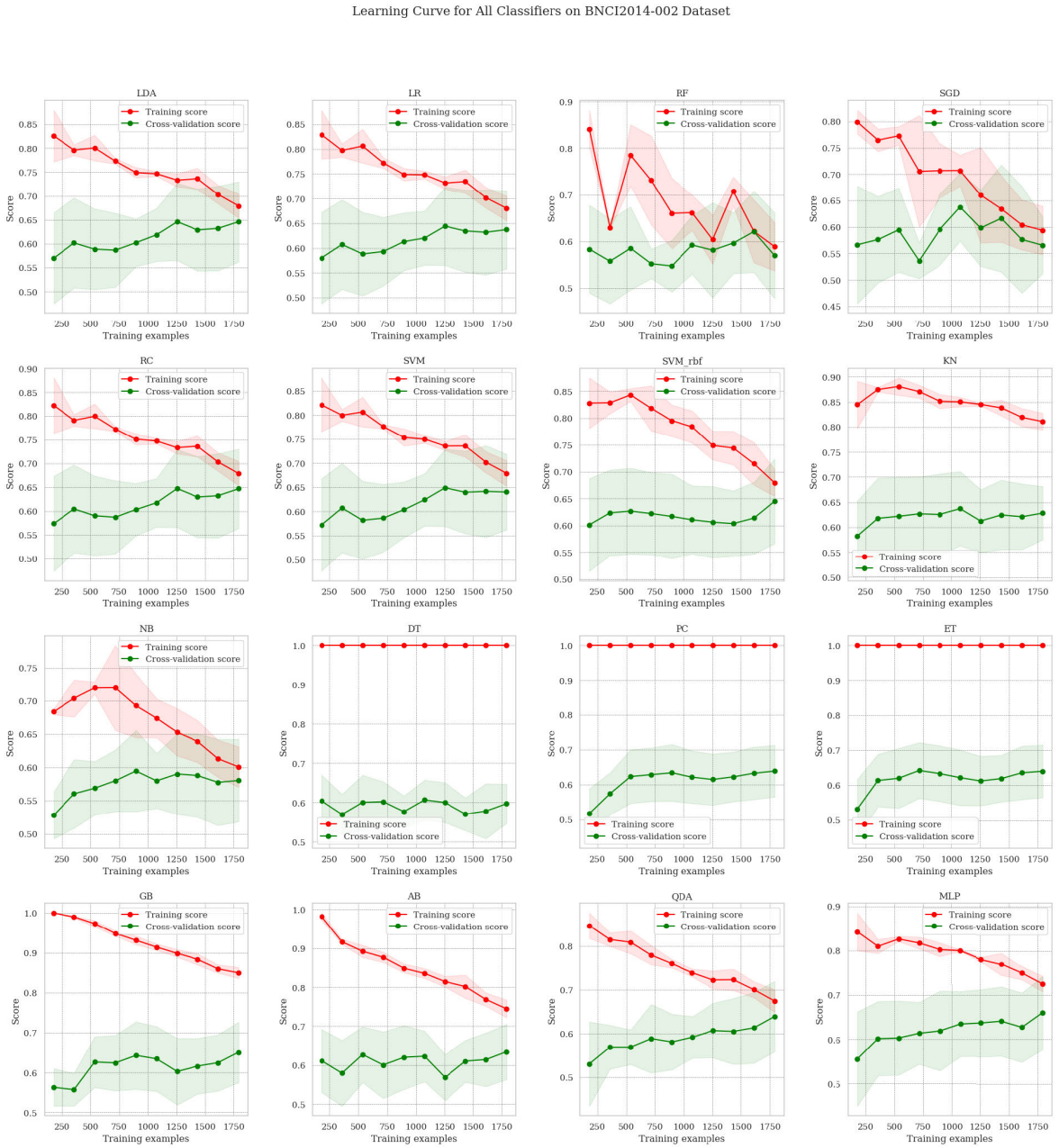

FIGURE 3. **Learning curves for all classifiers on the BNCI2014-002 dataset demonstrate the training and cross-validation score progression with increasing data volume. This plot provides insights into each classifier's learning efficiency and potential overfitting characteristics.**

interpretation of this analysis, illustrating how classifiers with similar performance profiles can be grouped to inform our ensemble model's structure.

Informed by the cluster analysis in Figure 4, we pinpoint the most effective classifiers within each identified cluster. Table 7 presents these classifiers along with their normalized weights, signifying their contribution to the diversity and efficacy of our proposed ensemble model.

Equipped with the insights gleaned from the cluster analysis, we constructed an ensemble model integrating the strengths of the best classifiers from each cluster. Finally, we highlight the outcomes of our comprehensive methodology, comparing the ensemble model's classification accuracy against those of the top-performing individual classifiers across all datasets. Figure 5 visually contrasts the accuracies, while Table 8 provides a detailed performance comparison. These results collectively demonstrate the ensemble model's superior ability to leverage the combined strengths of the classifiers, potentially enhancing prediction performance significantly.

TABLE 4. Learning curve analysis: Classifier performance across different datasets.

(a) BNCI2014-001

Classifier	AUC-CVR	CRR	PSR	AUC-CVN	CRN	PSN	Rank
LR	717	0.122	0.082	0.588	0.160	0.331	1
NB	681	0.066	0.066	0.839	0.057	0.000	2
PC	657	0.036	0.073	1.000	0.000	0.143	3
SVM	735	0.139	0.094	0.470	0.192	0.571	4
LDA	694	0.123	0.094	0.745	0.162	0.571	5
RC	692	0.112	0.099	0.760	0.142	0.673	6
MLP	800	0.139	0.107	0.023	0.192	0.837	7
SGD	668	0.133	0.085	0.926	0.180	0.388	8
AB	683	0.165	0.074	0.820	0.240	0.163	9
SVM-rbf	804	0.159	0.115	0.000	0.229	1.000	10
KN	772	0.237	0.101	0.217	0.373	0.714	11
QDA	724	0.143	0.105	0.540	0.199	0.796	12
GB	766	0.335	0.105	0.256	0.555	0.796	13
ET	794	0.485	0.108	0.063	0.833	0.857	14
RF	792	0.500	0.107	0.081	0.861	0.837	15
DT	681	0.575	0.088	0.837	1.000	0.449	16

(b) BNCI2014-002

Classifier	AUC-CVR	CRR	PSR	AUC-CVN	CRN	PSN	Rank
SVM-rbf	994	0.034	0.074	0.113	0.040	0.728	1
KN	1002	0.183	0.071	0.000	0.428	0.632	2
NB	931	0.020	0.053	1.000	0.004	0.044	3
GB	993	0.199	0.070	0.128	0.471	0.603	4
QDA	952	0.035	0.067	0.704	0.043	0.528	5
LR	993	0.043	0.077	0.132	0.063	0.855	6
SGD	949	0.028	0.077	0.741	0.024	0.826	7
PC	934	0.019	0.077	0.959	0.000	0.838	8
ET	998	0.363	0.076	0.063	0.900	0.804	9
RC	989	0.032	0.080	0.178	0.034	0.928	10
SVM	993	0.038	0.079	0.133	0.050	0.902	11
AB	980	0.111	0.072	0.317	0.239	0.686	12
MLP	1001	0.065	0.082	0.016	0.119	1.000	13
LDA	988	0.033	0.079	0.198	0.037	0.919	14
RF	990	0.360	0.074	0.152	0.892	0.736	15
DT	950	0.402	0.051	0.730	1.000	0.000	16

(c) BNCI2014-004

Classifier	AUC-CVR	CRR	PSR	AUC-CVN	CRN	PSN	Rank
LR	3206	0.013	0.053	0.000	0.011	0.982	1
LDA	3202	0.014	0.053	0.011	0.013	0.993	2
NB	3082	0.008	0.049	0.281	0.000	0.822	3
RC	3202	0.014	0.053	0.011	0.014	1.000	4
SVM	3191	0.017	0.051	0.036	0.023	0.925	5
SVM-rbf	3124	0.039	0.040	0.184	0.078	0.425	6
QDA	3160	0.017	0.051	0.106	0.022	0.891	7
MLP	3162	0.036	0.049	0.100	0.071	0.805	8
GB	3082	0.079	0.039	0.280	0.182	0.388	9
SGD	3063	0.020	0.046	0.323	0.029	0.687	10
RF	3005	0.351	0.034	0.453	0.886	0.168	11
AB	3076	0.047	0.046	0.294	0.099	0.674	12
KN	2982	0.137	0.038	0.505	0.333	0.336	13
ET	2999	0.352	0.037	0.467	0.889	0.277	14
DT	2818	0.395	0.030	0.889	1.000	0.000	15
PC	2761	0.038	0.049	1.000	0.076	0.799	16

(d) BNCI2015-001

Classifier	AUC-CVR	CRR	PSR	AUC-CVN	CRN	PSN	Rank
NB	2275	0.022	0.060	0.000	0.021	0.867	1
RC	2245	0.066	0.053	0.366	0.122	0.517	2
LDA	2245	0.066	0.053	0.361	0.123	0.531	3
LR	2256	0.072	0.054	0.227	0.138	0.570	4
SGD	2214	0.014	0.052	0.747	0.000	0.482	5
SVM	2232	0.091	0.045	0.527	0.182	0.186	6
RF	2266	0.431	0.052	0.114	0.980	0.485	7
AB	2239	0.173	0.045	0.433	0.375	0.169	8
SVM-rbf	2193	0.128	0.041	1.000	0.269	0.000	9
GB	2235	0.225	0.049	0.482	0.497	0.347	10
PC	2243	0.041	0.063	0.383	0.065	1.000	11
KN	2247	0.240	0.053	0.342	0.533	0.546	12
MLP	2255	0.160	0.058	0.244	0.345	0.778	13
QDA	2229	0.102	0.056	0.555	0.208	0.656	14
DT	2218	0.437	0.052	0.688	0.994	0.501	15
ET	2228	0.439	0.053	0.576	1.000	0.554	16

(e) Zhou2016

Classifier	AUC-CVR	CRR	PSR	AUC-CVN	CRN	PSN	Rank
SVM-rbf	907	0.067	0.051	0.000	0.146	0.060	1
RC	879	0.035	0.056	0.214	0.057	0.350	2
SVM	891	0.039	0.057	0.121	0.069	0.446	3
LDA	873	0.037	0.056	0.264	0.062	0.336	4
RF	896	0.277	0.051	0.085	0.734	0.026	5
LR	884	0.036	0.061	0.175	0.061	0.703	6
PC	827	0.014	0.055	0.615	0.000	0.304	7
AB	842	0.083	0.051	0.501	0.192	0.000	8
GB	880	0.191	0.052	0.211	0.491	0.098	9
MLP	886	0.062	0.059	0.161	0.134	0.568	10
ET	876	0.269	0.056	0.235	0.710	0.363	11
SGD	838	0.048	0.063	0.528	0.095	0.776	12
KN	829	0.151	0.057	0.604	0.380	0.427	13
QDA	839	0.052	0.066	0.524	0.104	1.000	14
NB	782	0.042	0.064	0.964	0.076	0.874	15
DT	777	0.373	0.063	1.000	1.000	0.811	16

(f) AlexMI

Classifier	AUC-CVR	CRR	PSR	AUC-CVN	CRN	PSN	Rank
SVM	251	0.154	0.112	0.213	0.172	0.506	1
KN	255	0.121	0.125	0.160	0.034	0.736	2
RC	267	0.112	0.139	0.000	0.000	1.000	3
NB	231	0.158	0.087	0.468	0.190	0.034	4
PC	257	0.200	0.124	0.128	0.362	0.724	5
LR	254	0.204	0.122	0.165	0.379	0.692	6
SVM-rbf	252	0.141	0.129	0.192	0.121	0.812	7
MLP	263	0.191	0.131	0.054	0.328	0.845	8
RF	252	0.316	0.116	0.189	0.845	0.569	9
AB	191	0.233	0.085	1.000	0.500	0.000	10
QDA	260	0.354	0.125	0.094	1.000	0.736	11
ET	258	0.300	0.129	0.112	0.776	0.813	12
LDA	249	0.166	0.131	0.231	0.224	0.842	13
GB	229	0.266	0.122	0.503	0.638	0.689	14
SGD	249	0.200	0.136	0.227	0.362	0.949	15
DT	209	0.316	0.124	0.762	0.845	0.732	16

Note: AUC-CVR = AUC-CV Raw, CRR = Convergence Rate Raw, PSR = Performance Stability Raw, AUC-CVN = AUC-CV Normalized, CRN = Convergence Rate Normalized, PSN = Performance Stability Normalized.

IV. DISCUSSION

This study presents the development and evaluation of the WS-AIEC, a novel approach for enhancing MI EEG signal classification. Our results demonstrate that the WS-AIEC significantly outperforms traditional classifiers, achieving remarkable accuracies of 96.88%, 99.58%, 96.25%, 98.75%, 94.58%, and 95.00% on the BNCI2014-001, BNCI2014-002, BNCI2014-004, BNCI2015-001, Zhou2016, and AlexMI datasets, respectively. These results highlight the effectiveness of the WS-AIEC model in dealing with the complexities and variabilities inherent in EEG data, particularly in MI tasks.

TABLE 5. Consolidated learning curve rankings of classifiers across all datasets.

Classifier	NB	PC	SGD	RC	SVM	AB	LDA	SVM-rbf	LR	QDA	MLP	GB	KN	DT	RF	ET
Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

TABLE 6. Overall ranking of classifiers based on composite scores with corresponding weights for ensemble integration.

Classifier	SVM	RC	SVM-rbf	LR	NB	SGD	MLP	PC	LDA	KN	AB	QDA	GB	RF	ET	DT
Overall Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Weight	0.296	0.148	0.099	0.074	0.059	0.049	0.042	0.037	0.033	0.030	0.027	0.025	0.023	0.021	0.020	0.018

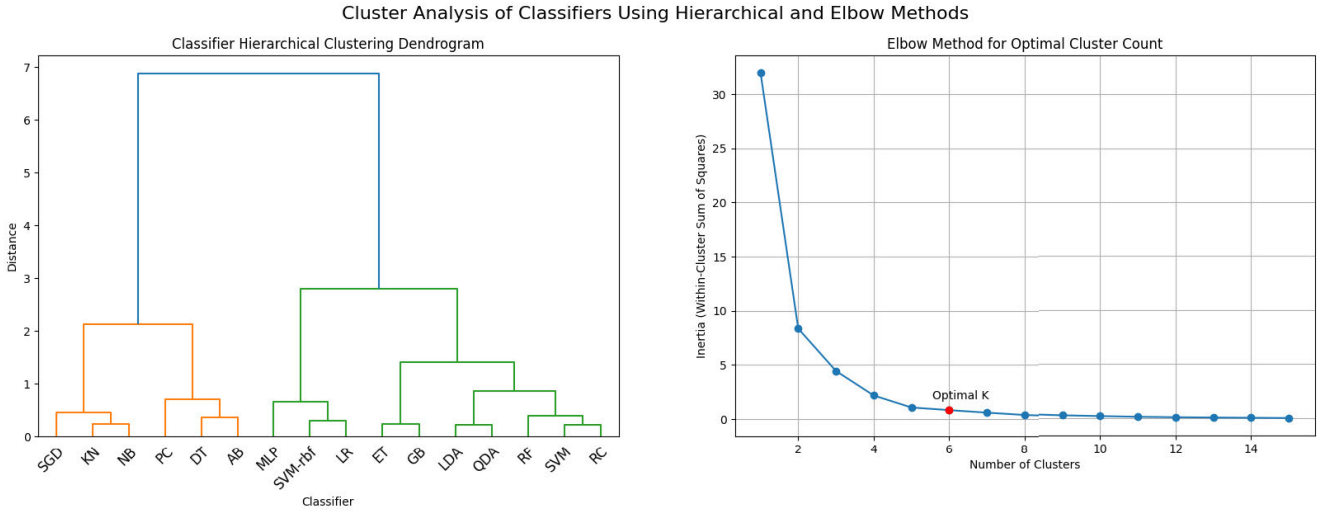

FIGURE 4. Cluster analysis of classifiers using hierarchical clustering (left) and the elbow method (right) to determine the optimal number of clusters for ensemble model selection.

TABLE 7. Best classifier in each cluster with corresponding classifiers and normalised weights.

Cluster	Classifiers	Best Classifier	Overall Rank	Normalised Weight
1	SGD, KN, NB	NB	5	0.1079
2	PC, DT, AB	PC	8	0.0676
3	MLP, SVM-rbf, LR	SVM-rbf	3	0.1810
4	ET, GB	GB	13	0.0420
5	LDA, QDA	LDA	9	0.0603
6	RF, SVM, RC	SVM	1	0.5411

After demonstrating the ensemble model's superior performance in classifying MI EEG signals with unprecedented accuracy levels, we must also acknowledge the parallel advancements in signal processing techniques that enhance our approach. Notably, a novel non-linear spatiotemporal filtering technique introduced recently aims to boost the detection accuracy of movement-related cortical potentials in single-trial EEG data [33]. This approach significantly contributes to the field by enhancing the detection accuracy of cortical potentials, showcasing the capacity of innovative signal processing methods in improving BCI systems. Such advancements highlight the importance of integrating sophisticated signal processing algorithms with ensemble learning approaches to expand the possibilities of what BCI technology can achieve.

A. DIRECT PERFORMANCE COMPARISON OF CLASSIFIERS

As detailed in Table 2, our evaluation method centers on the accuracy, precision, recall, F1 score, AUC-ROC, and kappa to ascertain the most proficient classifiers for our ensemble model.

This approach highlighted the standout performance of classifiers like SVM, demonstrating the highest overall scores across our datasets. These findings confirm that SVM and other high-performing classifiers offer considerable advancements in managing the complexities of EEG signals.

The performance rankings summarized in Table 3 provide a holistic perspective on classifier efficacy, ensuring a balanced and data-driven foundation for the WS-AIEC model. Our ensemble thus reflects contemporary classifier capabilities, strategically combined to amplify their strengths. The results from this direct performance comparison guide the construction of an ensemble model tailored to excel in accuracy and reliability.

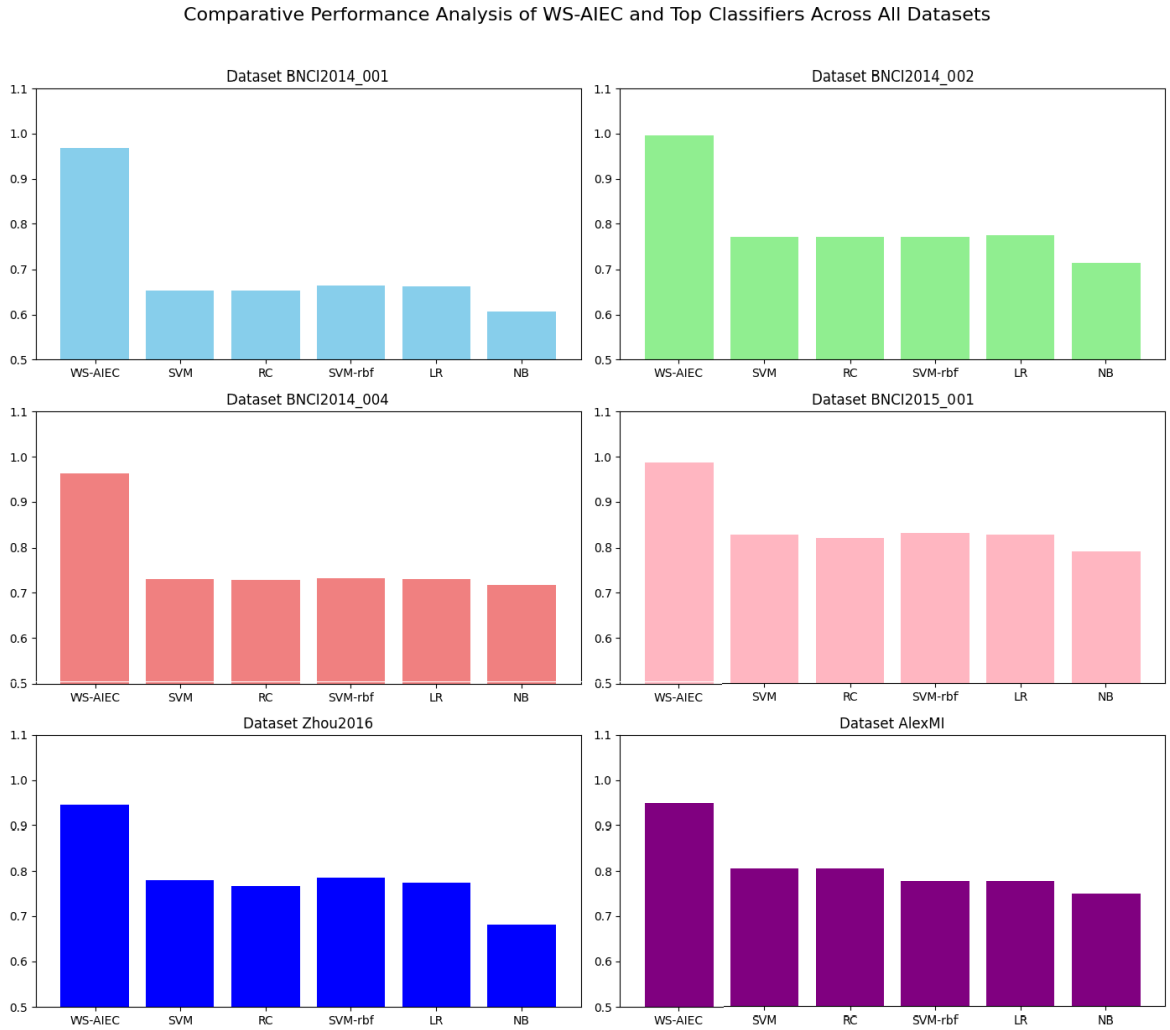

FIGURE 5. Classification accuracies of the WS-AIEC model versus the top five individual classifiers across six datasets. Each bar plot contrasts the accuracy of the ensemble model with the individual performances, highlighting the ensemble's relative effectiveness.

B. INSIGHTS FROM LEARNING CURVE ANALYSIS

In Figure 3, we showcased the learning curves for all classifiers on the BNCI2014-002 dataset. The selection of this dataset as a representative example was due to its general reflection of trends observed across other datasets. The figure captures how classifier performance improves with more data and reveals the point at which further training ceases to yield benefits—a phenomenon of overfitting.

The quantified learning curve data in Table 4 complements this visual assessment. We can conclude each classifier's reliability and predictive strength by ranking classifiers based on key metrics such as AUC-CV, CR, and PS. Notably, classifiers with a low convergence rate across training iterations demonstrate a robust learning strategy that is less prone to the pitfalls of overfitting and better suited for generalizing to new data.

This analysis brought to light several key findings. The NB classifier, for instance, despite not consistently ranking high in individual dataset performance, showcased remarkable learning stability and generalization as evidenced by its high rank in the overall learning curve analysis. Conversely, classifiers like DT, which typically demonstrate rapid early learning, ranked lower due to less stability in performance as more data was introduced.

Furthermore, the consolidated learning curve rankings presented in Table 5 provide a singular, comprehensive ranking system across all datasets, reinforcing the importance of generalization in classifier evaluation. This ranking underscores classifiers' adaptability to the multifaceted nature of EEG data, which is often riddled with noise and inter-subject variability.

C. OPTIMISING ENSEMBLE STRUCTURE THROUGH CLASSIFIER CLUSTERING

The clustering analysis, as depicted in Figure 4, represents an advanced stage in our ensemble model's construction. Employing hierarchical clustering and the elbow method, we identified natural groupings among classifiers based on

TABLE 8. Performance comparison of WS-AIEC and top classifiers.

(a) BNCI2014-001

Classifier	Accuracy	Precision	Recall	F1 Score	AUC-ROC	Kappa
WS-AIEC	**0.9688**	**0.9700**	**0.9688**	**0.9687**	**0.9688**	**0.9375**
SVM	0.6542	0.6618	0.6542	0.6536	0.7687	0.5389
RC	0.6542	0.6560	0.6542	0.6506	0.7679	0.5378
SVM-rbf	0.6638	0.6709	0.6638	0.6623	0.7752	0.5516
LR	0.6619	0.6694	0.6619	0.6619	0.7739	0.5489
NB	0.6073	0.6105	0.6073	0.6025	0.7360	0.4755

(b) BNCI2014-002

Classifier	Accuracy	Precision	Recall	F1 Score	AUC-ROC	Kappa
WS-AIEC	**0.9958**	**0.9959**	**0.9958**	**0.9958**	**0.9972**	**0.9944**
SVM	0.7723	0.7843	0.7723	0.7720	0.7754	0.5436
RC	0.7723	0.7846	0.7723	0.7724	0.7758	0.5440
SVM-rbf	0.7723	0.7861	0.7723	0.7719	0.7765	0.5449
LR	0.7768	0.7852	0.7768	0.7768	0.7796	0.5515
NB	0.7143	0.7200	0.7143	0.7141	0.7153	0.4257

(c) BNCI2014-004

Classifier	Accuracy	Precision	Recall	F1 Score	AUC-ROC	Kappa
WS-AIEC	**0.9625**	**0.9646**	**0.9625**	**0.9624**	**0.9625**	**0.9250**
SVM	0.7304	0.7364	0.7304	0.7291	0.7322	0.4624
RC	0.7290	0.7360	0.7290	0.7282	0.7314	0.4605
SVM-rbf	0.7331	0.7400	0.7331	0.7312	0.7345	0.4677
LR	0.7297	0.7375	0.7297	0.7284	0.7329	0.4631
NB	0.7173	0.7226	0.7173	0.7161	0.7183	0.4348

(d) BNCI2015-001

Classifier	Accuracy	Precision	Recall	F1 Score	AUC-ROC	Kappa
WS-AIEC	**0.9875**	**0.9879**	**0.9875**	**0.9875**	**0.9891**	**0.9745**
SVM	0.8274	0.8319	0.8274	0.8271	0.8287	0.6551
RC	0.8191	0.8216	0.8191	0.8190	0.8195	0.6379
SVM-rbf	0.8312	0.8361	0.8312	0.8309	0.8325	0.6627
LR	0.8267	0.8301	0.8267	0.8266	0.8277	0.6534
NB	0.7917	0.8081	0.7917	0.7839	0.7920	0.5826

(e) Zhou2016

Classifier	Accuracy	Precision	Recall	F1 Score	AUC-ROC	Kappa
WS-AIEC	**0.9458**	**0.9474**	**0.9458**	**0.9458**	**0.9462**	**0.8912**
SVM	0.7795	0.7856	0.7795	0.7788	0.8361	0.6695
RC	0.7672	0.7698	0.7672	0.7654	0.8269	0.6509
SVM-rbf	0.7860	0.7946	0.7860	0.7849	0.8408	0.6792
LR	0.7745	0.7788	0.7745	0.7737	0.8319	0.6617
NB	0.6805	0.6882	0.6805	0.6745	0.7589	0.5186

(f) AlexMI

Classifier	Accuracy	Precision	Recall	F1 Score	AUC-ROC	Kappa
WS-AIEC	**0.9500**	**0.9517**	**0.9500**	**0.9499**	**0.9492**	**0.8995**
SVM	0.8056	0.8727	0.8056	0.8182	0.8489	0.7040
RC	0.8056	0.8389	0.8056	0.8025	0.8494	0.6998
SVM-rbf	0.7778	0.8505	0.7778	0.7837	0.8327	0.6649
LR	0.7770	0.8217	0.7770	0.7764	0.8321	0.6575
NB	0.7500	0.7847	0.7500	0.7601	0.7938	0.6111

their performance characteristics. This step was crucial for understanding the relationships and redundancies between different classifiers, allowing us to harness diversity effectively while avoiding duplication in the ensemble's decision-making process.

Table 7 elucidates each identified cluster's best-performing classifier and normalized weights. Notably, the clustering analysis ensures that our ensemble model capitalizes on the collective strengths of classifiers from each group, enhancing overall performance through a strategic blend of varied expertise.

D. RATIONALE FOR SVM AS THE META-CLASSIFIER

Upon consolidating the rankings and clustering insights, selecting a meta-classifier became pivotal. SVM emerged as the preferred choice, and for good reasons, elucidated throughout our discussion. This decision is a product of its performance and its compatibility with the ensemble.

The choice of SVM is justified by its demonstrated balance between complexity and generalization, making it ideal for synthesizing inputs from diverse classifiers without succumbing to overfitting. Its success in high-dimensional spaces is particularly salient for the ensemble model, which must interpret a complex input array from base classifiers.

Moreover, the linear version of SVM offers a compromise between capturing complex relationships and preserving interpretability, a balance that is often challenging to achieve. By opting for SVM, we ensure that our ensemble model benefits from improved accuracy and robustness while maintaining generalizability.

E. ADAPTIVE WEIGHTING: ENHANCING MODEL PERFORMANCE THROUGH DYNAMIC ADJUSTMENT

The rationale behind dynamic weight adjustment stems from the understanding that data characteristics can change over time or across different data segments. Traditional static weighting approaches do not account for this variability, potentially leading to suboptimal ensemble performance. By introducing a mechanism for dynamic adjustment, our model can better adapt to such changes, thereby improving its generalization ability and performance on unseen data. This dynamic approach allows the ensemble to recalibrate the influence of each classifier based on real-time performance metrics across varied conditions. We validate this adaptability through rigorous testing on multiple datasets, ensuring robust performance across different time segments and under different operational conditions. This process is crucial for handling the inherent non-stationarity of EEG signals and the evolving nature of BCI tasks.

F. COMPARATIVE PERFORMANCE AND FINAL VALIDATION OF THE WS-AIEC MODEL

The conclusive phase of our research presents a compelling comparison of the WS-AIEC ensemble model against the top individual classifiers, as illustrated in Figure 5. This juxtaposition not only demonstrates the superiority of the

ensemble approach but also validates the efficacy of our comprehensive classifier integration strategy.

The bar plots in Figure 5 succinctly contrast the classification accuracies of the ensemble model with those of the individual classifiers. The WS-AIEC model consistently outperforms each classifier, showcasing its robustness and adaptability across all datasets.

The accompanying Table 8 further quantifies the WS-AIEC model's performance, providing a granular view of its precision, recall, F1 score, AUC-ROC, and Kappa statistics. These metrics solidify the WS-AIEC's status as a high-performing model characterized by balance and precision across various evaluation dimensions.

G. ABLATION ANALYSIS

To understand the contribution of each component in our methodology, we conducted a series of ablation studies by systematically modifying the selection criteria for classifiers and evaluating the impact on model performance, as illustrated in Figure 1. We built the ensemble for each scenario and tested it on two datasets: BNCI2014-002 and Zhou2016. These datasets were selected because the original ensemble model achieved the highest accuracy of 0.9958 on BNCI2014-002 and the lowest accuracy of 0.9458 on Zhou2016. The results for each scenario are presented in Table 9 and summarized below. The ablation analysis reveals several key insights into the importance and interaction of different metrics and classifiers within the ensemble model.

Firstly, Scenario 1, which utilizes only Comprehensive Performance Assessment (CPA), shows the most significant changes in the selected classifiers, indicating that Learning Curve Analysis (LCA) is crucial in optimizing the ensemble. In Scenario 2, where only LCA is used, the performance drop is more pronounced despite having fewer changes in classifiers (3 changes) compared to Scenario 1 (4 changes). This suggests that CPA is more critical than LCA, as the absence of CPA leads to a greater decrease in accuracy even with fewer classifier changes.

In Scenarios 3, 4, and 5, where one LCA metric is removed each time, we observe that each scenario results in only one classifier change. The impact on performance varies significantly based on the specific metric removed. Scenario 3, which replaces GB with RF (overall rank 14), shows minimal performance reduction. In contrast, Scenarios 4 and 5, which replace LDA with RC (overall rank 2) and GB with SGD (overall rank 6), exhibit greater performance drops. This highlights that the difference in overall ranking between the replaced classifiers directly affects performance, with larger differences resulting in more significant accuracy reductions.

For Scenarios 6, 7, and 8, where only one LCA metric is used, the number of classifier changes increases compared to Scenarios 3, 4, and 5. Scenario 7, with only two changes, shows a smaller performance decrease than Scenarios 6 and 8, which have three changes each. Notably, the presence of CPA in Scenarios 6 and 8 mitigates the performance decrease compared to Scenario 2 despite having the same number of classifier changes. This underscores the essential role of CPA in maintaining ensemble performance.

A closer look at Scenarios 6 and 8, where three changes occur, reveals that two (SGD and MLP replacing LDA and GB) are consistent between the scenarios. The third change differs from Scenario 6, which replaces PC with AB (overall rank 11), while Scenario 8 replaces SVM-rbf with RC (overall rank 2). The slightly higher performance drop in Scenario 6 corresponds to the greater difference in ranking between the replaced classifiers, illustrating the importance of classifier ranking in the ensemble.

Two experiments were conducted in Scenarios 9 and 10 to investigate individual classifiers' effectiveness further. First, removing GB in Scenario 9, the lowest-ranked performer, resulted in a 1.79% accuracy decrease on BNCI2014-002 and 1.56% on Zhou2016, indicating that even lower-ranked classifiers significantly contribute to the ensemble. Second, removing SVM-rbf (Scenario 10), one of the top performers, led to a much larger accuracy drop, resulting in a 7.78% accuracy decrease on BNCI2014-002 and 7.32% on Zhou2016, highlighting the critical role of higher-ranked classifiers in the ensemble's overall performance. Notably, the ratios of the accuracy decrease—$(7.78/1.79) \approx 4.35$ and $(7.32/1.56) \approx 4.69$—closely align with the ratio of their weights in Table 7 ($0.1810/0.0420 \approx 4.30$). This further validates the effectiveness and significance of our algorithm in selecting the most appropriate classifiers for the ensemble model.

Finally, the ensemble's performance with static weights was tested in Scenario 11. Using static weights based on overall ranks instead of dynamic weights led to a notable decrease in accuracy, demonstrating the effectiveness of the dynamic weight assignment algorithm in adapting to varying data characteristics and maintaining high performance.

As depicted in Figure 6, the performance impact of different ablation scenarios is visually represented, highlighting the variations in accuracy across the BNCI2014-002 and Zhou2016 datasets.

This detailed ablation analysis demonstrates the robustness of our methodology and the significance of each component in achieving optimal performance. Removing individual classifiers like GB or SVM-rbf significantly impacts accuracy, highlighting their importance in the ensemble. Dynamic weighting also plays a critical role, allowing the model to adapt to changes in classifier performance and ensuring consistently high accuracy. Overall, the WS-AIEC model's ability to integrate diverse classifiers and dynamically adjust their contributions based on real-time performance evaluations sets it apart, achieving superior classification performance across different datasets.

H. COMPARISON WITH RELATED WORKS

The WS-AIEC model's performance is further highlighted compared to the performances reported in related studies. Table 10 provides a comprehensive comparison, showcasing

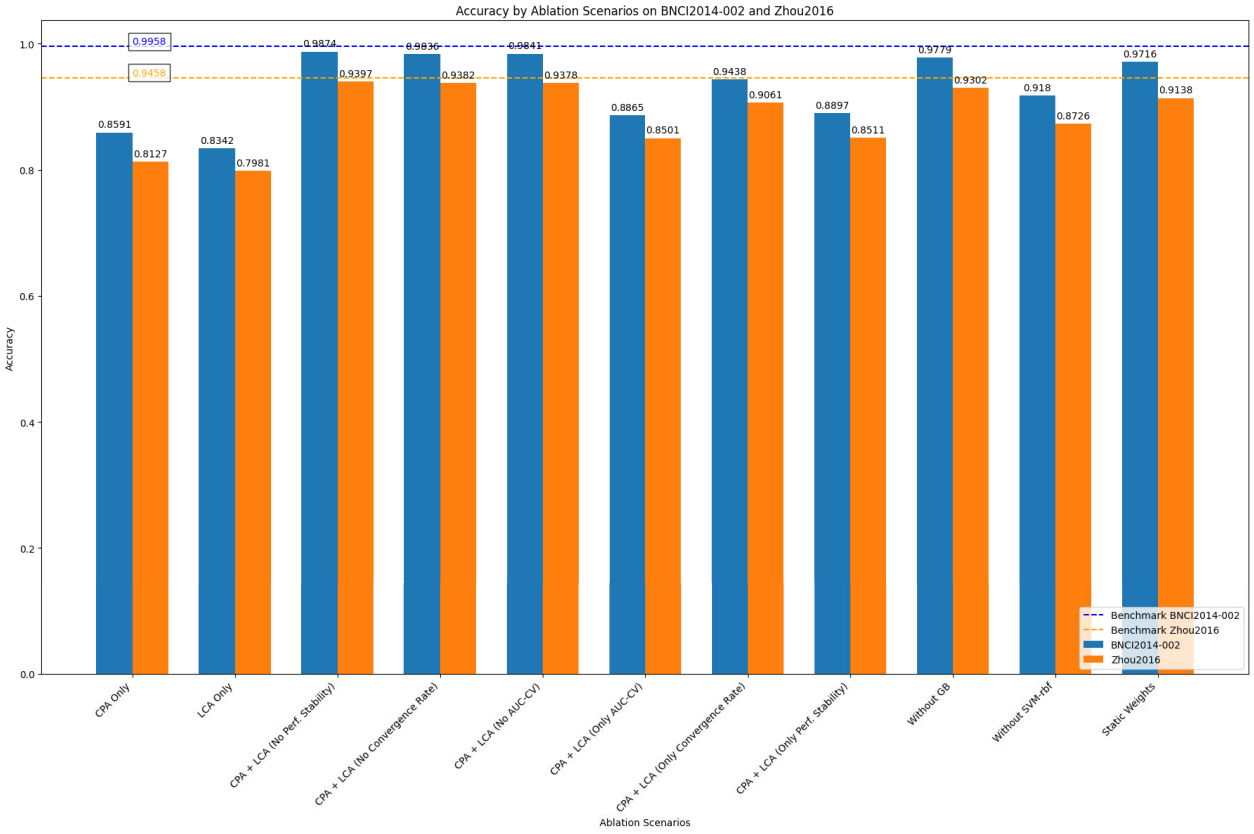

FIGURE 6. Accuracy by ablation scenarios on BNCI2014-002 and Zhou2016 datasets. The figure compares the performance of the ensemble model across various ablation scenarios, highlighting the impact of different metrics and classifiers on the overall accuracy. The blue bars represent accuracy on the BNCI2014-002 dataset, while the orange bars represent accuracy on the Zhou2016 dataset. Dashed lines for each dataset indicate benchmark accuracies.

TABLE 9. Ablation analysis results.

Scenario	Selected Classifiers	Accuracy on BNCI2014-002	Accuracy on Zhou2016
Scenario 1: CPA Only	SVM, SVM-rbf, MLP, LR, RC, KN	0.8591	0.8127
Scenario 2: LCA Only	NB, PC, SGD, RC, SVM, AB	0.8342	0.7981
Scenario 3: CPA + LCA (No Performance Stability)	NB, PC, SVM-rbf, RF, LDA, SVM	0.9874	0.9397
Scenario 4: CPA + LCA (No Convergence Rate)	NB, PC, SVM-rbf, GB, RC, SVM	0.9836	0.9382
Scenario 5: CPA + LCA (No AUC-CV)	SVM, LDA, SVM-rbf , PC, NB, SGD	0.9841	0.9378
Scenario 6: CPA + LCA (Only AUC-CV)	NB, MLP, SVM-rbf, AB, SGD, SVM	0.8865	0.8501
Scenario 7: CPA + LCA (Only Convergence Rate)	NB, PC, RF, SGD, SVM-rbf, SVM	0.9438	0.9061
Scenario 8: CPA + LCA (Only Performance Stability)	MLP, SVM, RC, PC, NB, SGD	0.8897	0.8511
Scenario 9: Without GB	NB, PC, SVM-rbf, LDA, SVM	0.9779	0.9302
Scenario 10: Without SVM-rbf	NB, PC, GB, LDA, SVM	0.9180	0.8726
Scenario 11: Static Weights	NB, PC, SVM-rbf, GB, LDA, SVM	0.9716	0.9138

the WS-AIEC model's superior performance metrics, which indicate its robustness and advanced classification capabilities. Notably, our model achieves a high classification accuracy and exhibits proficiency across multiple datasets, indicating its strong generalization abilities.

In contrast with existing methodologies, the WS-AIEC model integrates a multitude of classifiers and employs a unique combination of weighted and stacking approaches. These strategies are underscored by our adaptive methodology, which dynamically adjusts classifier weights in response to data variability, a feature not commonly reported in the literature. Clustering or segmentation techniques refine our ensemble model's structure, optimizing decision-making by strategically selecting classifiers.

Particularly noteworthy is the comparison with our previous work [20], which utilized a significant number of classifiers across multiple datasets. Our current WS-AIEC model surpasses the accuracy benchmark set by that study and demonstrates a marked advancement in ensemble approaches. On average, the WS-AIEC model exhibits an accuracy improvement of approximately 1.47%, further emphasizing the refined methodology's impact. Moreover, even in the worst-case scenario, the WS-AIEC model shows a minimum improvement of 0.87% in accuracy,

TABLE 10. Performance comparison of WS-AIEC model with related works.

Refrence	Classifiers	Datasets	EME[1]	WAU[2]	SAU[3]	AME[4]	CST[5]	Accuracy(%)
Nicolas et al., 2014 [34]	1	1	✓	-	✓	-	-	92.00
Rahimi et al., 2016 [35]	3	1	✓	-	✓	-	-	90.27
Mohammadpour et al., 2016 [36]	3	1	✓	-	-	-	-	73.42
Datta et al., 2017 [37]	2	1	✓	-	-	-	-	83.57
Ramos et al., 2017 [38]	3	2	✓	✓	-	✓	-	97.48
Chatterjee et al., 2018 [39]	5	1	✓	-	-	-	-	85.71
Raza et al., 2019 [40]	1	2	✓	✓	-	✓	-	81.48
Salimi et al., 2019 [14]	1	1	✓	-	-	✓	-	96.00
Zhang et al., 2020 [41]	1	2	✓	-	-	✓	-	90.12
Tyagi et al., 2020 [12]	3	1	✓	✓	-	✓	-	85.83
Norizadeh et al., 2021 [42]	5	2	✓	-	-	✓	-	86.91
Nugroho et al., 2021 [43]	2	1	✓	-	-	-	-	83.28
Rashid et al., 2021 [44]	4	4	✓	-	-	-	-	99.21
Wei et al., 2021 [15]	5	2	✓	-	-	-	-	82.99
Zheng et al., 2021 [45]	1	1	✓	✓	-	-	-	81.58
Sun et al., 2021 [46]	5	1	✓	-	✓	-	-	87.80
Du et al., 2021 [18]	1	1	✓	✓	-	✓	-	95.00
Kamhi et al., 2022 [47]	5	1	✓	-	-	-	-	92.00
Dolzhikova et al., 2022 [17]	1	1	✓	✓	-	✓	-	95.58
Quanyu et al., 2023 [48]	4	1	✓	-	-	-	-	91.46
Mehtiyev et al., 2023 [11]	4	1	✓	✓	-	-	-	96.98
Almohammadi et al., 2023 [13]	15	1	✓	-	✓	-	-	86.23
Esfahani et al., 2023 [16]	3	2	✓	✓	-	-	-	97.50
Shin et al., 2024 [49]	1	1	✓	-	✓	✓	-	91.71
Hossein et al., 2024 [20]	**16**	4	✓	✓	✓	✓	-	98.16
This Study	**16**	**6**	✓	✓	✓	✓	✓	**99.58**

[1]**EME**: Ensemble Methodology Employed
[2]**WAU**: Weighted Approach Used
[3]**SAU**: Stacking Approach Used
[4]**AME**: Adaptive Methodology Employed
[5]**CST**: Clustering or Segmentation Techniques

underscoring the consistent enhancements achieved through our methodical and adaptive ensemble strategies.

The breadth of classifiers used in the WS-AIEC model, combined with our comprehensive ensemble strategy, represents an evolution in MI EEG classification. This progression resonates with the increasing complexity and variety within EEG datasets, demanding more sophisticated and adaptive solutions. Our ensemble model does not simply rely on a large number of classifiers but utilizes a methodical integration and dynamic weighting system that facilitates a high degree of predictive accuracy.

The outcomes highlighted in Table 10 underscore the superior efficacy of the WS-AIEC model, which not only improves upon other singular approaches but also advances beyond our previous research. Despite the variety in datasets used across the studies we compare, our model demonstrates exceptional adaptability and robustness, consistently surpassing the performance benchmarks of earlier works under equivalent conditions. This adaptability highlights the WS-AIEC model's innovative capabilities within adaptive ensemble approaches, setting a new benchmark for EEG classification. It is important to note that our comparisons focus on the highest accuracy achieved by each method, providing a clear benchmark for assessing the peak performance and offering a focused comparison of the most successful outcomes.

I. LIMITATIONS AND FUTURE DIRECTIONS

While the WS-AIEC model exhibits exceptional classification accuracy and robustness, it also presents certain limitations. The computational complexity, primarily due to integrating multiple classifiers and the intricate weighting mechanism, is a pertinent challenge. Future research could explore optimization strategies to enhance computational efficiency without sacrificing accuracy.

Additionally, the WS-AIEC's real-world applicability, particularly in complex and noisy data scenarios, necessitates further investigation. Extending its application to real-time BCI systems and assessing its adaptability to individual user characteristics are crucial future steps.

Building on the success of the WS-AIEC model, future research avenues are vast. They include exploring the impacts of different data preprocessing techniques, the integration of additional classifiers, or applying ensemble models to other complex classification tasks. The model's foundations in adaptive learning and model optimization set the stage for these endeavors, providing a robust starting point for future innovations in dynamic ensemble methods. Future work may delve into more sophisticated weight adjustment algorithms, evaluate the consequences of varying evaluation subset sizes and selection criteria, and incorporate additional performance metrics into the weight adjustment process to cultivate a more versatile and responsive ensemble framework.

While this study focuses on developing an ensemble model for MI EEG classification using CSP for feature extraction, future work could explore integrating various advanced feature extraction techniques. This includes phase-based approaches, instantaneous phase difference sequences, divergence-based features, and other innovative methods mentioned in recent studies [50]. Evaluating these techniques within our ensemble framework could potentially enhance classification performance and provide deeper insights into MI EEG signal processing.

Furthermore, another promising direction is incorporating learning-based approaches, such as the Hamilton-Jacobi-Bellman (HJB) equation-based learning scheme for neural networks. This method could provide a more performing framework for ensemble learning by leveraging advanced neural network optimization techniques, potentially leading to significant improvements in classification accuracy and model robustness [51].

Future research should also validate the WS-AIEC model using in vivo datasets to strengthen its practical implications and reliability. This involves testing the model with data collected from live subjects under operational conditions, demonstrating its applicability and robustness in real-world scenarios, and enhancing its relevance and potential impact in practical BCI applications.

V. CONCLUSION

In conclusion, the WS-AIEC model represents a significant leap forward in MI EEG signal classification, offering a robust and reliable tool that stands out for its accuracy and adaptability. This study achieves superior performance by judiciously harnessing the complementary strengths of diverse classifiers and laying the groundwork for more nuanced and sophisticated ensemble approaches. Looking ahead, the methodologies and insights garnered from this research can be adapted to a range of complex systems, inspiring future innovations in BCI technology and beyond. Our findings advocate for the continued exploration of ensemble learning strategies, reinforcing the value of comprehensive evaluation frameworks in pursuing breakthroughs in EEG signal processing.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan, "Brain–computer interface technology: A review of the first international meeting," in *Proc. IEEE Trans. Rehabil. Eng.*, Jun. 2000, vol. 8, no. 2, pp. 164–173, doi: 10.1109/TRE.2000.847807.

[2] S. Makeig, C. Kothe, T. Mullen, N. Bigdely-Shamlo, Z. Zhang, and K. Kreutz-Delgado, "Evolving signal processing for brain–computer interfaces," *Proc. IEEE*, vol. 100, pp. 1567–1584, May 2012, doi: 10.1109/JPROC.2012.2185009.

[3] D. L. Schomer and F. H. Lopes da Silva, *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 6th ed., Philadelphia, PA, USA: Wolters Kluwer Health, 2011.

[4] L. Mesin, G. E. Cipriani, and M. Amanzio, "Electroencephalography-based brain–machine interfaces in older adults: A literature review," *Bioengineering*, vol. 10, no. 4, p. 395, Mar. 2023, doi: 10.3390/bioengineering10040395.

[5] J. Meng, S. Zhang, A. Bekyo, J. Olsoe, B. Baxter, and B. He, "Noninvasive electroencephalogram based control of a robotic arm for reach and grasp tasks," *Sci. Rep.*, vol. 6, no. 1, Dec. 2016, Art. no. 38565, doi: 10.1038/srep38565.

[6] C. M. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics). New York, NY, USA: Springer, 2006. [Online]. Available: https://link.springer.com/book/9780387310732

[7] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 4, no. 2, pp. R1–R13, Jun. 2007, doi: 10.1088/1741-2560/4/2/r01.

[8] T. G. Dietterich, "Ensemble methods in machine learning," in *Proc. Int. Workshop Multiple Classifier Syst.* Berlin, Heidelberg: Springer, 2000, pp. 1–15.

[9] L. Rokach, "Ensemble-based classifiers," *Artif. Intell. Rev.*, vol. 33, nos. 1–2, pp. 1–39, Feb. 2010, doi: 10.1007/s10462-009-9124-7.

[10] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*, 1st ed., Boca Raton, FL, USA: CRC Press, 2012, doi: 10.1201/b12207.

[11] A. Mehtiyev, A. Al-Najjar, H. Sadreazami, and M. Amini, "DeepEnsemble: A novel brain wave classification in MI-BCI using ensemble of deep learners," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Las Vegas, NV, USA, Jan. 2023, pp. 1–5, doi: 10.1109/ICCE56470.2023.10043385.

[12] A. Tyagi and S. Semwal, "Ensemble classifiers for brain cyborgs," in *Proc. Int. Conf. Adv. Comput., Commun. Mater. (ICACCM)*, Dehradun, India, Aug. 2020, pp. 18–22, doi: 10.1109/ICACCM50413.2020.9213054.

[13] A. Almohammadi and Y.-K. Wang, "Integrated connectivity-based stacking ensemble learning with GCNNs for EEG representation," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Mexico City, Mexico, Dec. 2023, pp. 41–46, doi: 10.1109/ssci52147.2023.10371894.

[14] N. Salimi, M. Barlow, and E. Lakshika, "Mental workload classification using short duration EEG data: An ensemble approach based on individual channels," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Xiamen, China, Dec. 2019, pp. 393–398, doi: 10.1109/SSCI44817.2019.9003141.

[15] X. Wei, E. Dong, and L. Zhu, "Multi-class MI-EEG classification: Using FBCSP and ensemble learning based on majority voting," in *Proc. China Autom. Congr. (CAC)*, Beijing, China, Oct. 2021, pp. 872–876, doi: 10.1109/CAC53003.2021.9728576.

[16] M. M. Esfahani, M. H. Najafi, and H. Sadati, "Optimizing EEG signal classification for motor imagery BCIs: FilterBank CSP with Riemannian manifolds and ensemble learning models," in *Proc. 9th Int. Conf. Signal Process. Intell. Syst. (ICSPIS)*, Bali, Bali, Indonesia, Dec. 2023, pp. 1–6, doi: 10.1109/icspis59665.2023.10402664.

[17] I. Dolzhikova, B. Abibullaev, R. Sameni, and A. Zollanvari, "Subject-independent classification of motor imagery tasks in EEG using multisubject ensemble CNN," *IEEE Access*, vol. 10, pp. 81355–81363, 2022, doi: 10.1109/ACCESS.2022.3195513.

[18] C. Du, C. Shi, H. Huang, and X. Wu, "The motor imagery EEG classification method combining common spatial pattern and ensemble learning," in *Proc. 6th Int. Conf. Commun., Image Signal Process. (CCISP)*, Chengdu, China, Nov. 2021, pp. 361–366, doi: 10.1109/CCISP52774.2021.9639289.

[19] M.-C. Corsi, S. Chevallier, F. D. V. Fallani, and F. Yger, "Functional connectivity ensemble method to enhance BCI performance (FUCONE)," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 9, pp. 2826–2838, Sep. 2022, doi: 10.1109/TBME.2022.3154885.

[20] H. Ahmadi and L. Mesin, "Enhancing motor imagery electroencephalography classification with a correlation-optimized weighted stacking ensemble model," *Electronics*, vol. 13, no. 6, p. 1033, Mar. 2024, doi: 10.3390/electronics13061033.

[21] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017, doi: 10.1002/hbm.23730.

[22] R. O. Duda, P. E. Hart, and D. G. Stork. (2019). *Pattern Classification*. Accessed: Jul. 6, 2023. [Online]. Available: https://lmsasu.github.io/MLReadingGroup/Presentations/2019/01.January/22/.pdf

[23] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, p. 55, 2012, doi: 10.3389/fnins.2012.00055. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2012.00055/full

[24] D. Steyrl, R. Scherer, J. Faller, and G. R. Müller-Putz, "Random forests in non-invasive sensorimotor rhythm brain–computer interfaces: A practical and convenient non-linear classifier," *Biomed. Eng./Biomedizinische Technik*, vol. 61, no. 1, pp. 77–86, Feb. 2016, doi: 10.1515/bmt-2014-0117.

[25] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, "Brain–computer communication: Motivation, aim, and impact of exploring a virtual apartment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 4, pp. 473–482, Dec. 2007, doi: 10.1109/tnsre.2007.906956.
[26] J. Faller, C. Vidaurre, T. Solis-Escalante, C. Neuper, and R. Scherer, "Autocalibration and recurrent adaptation: Towards a plug and play online ERD-BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 3, pp. 313–319, May 2012, doi: 10.1109/TNSRE.2012.2189584.
[27] B. Zhou, X. Wu, Z. Lv, L. Zhang, and X. Guo, "A fully automated trial selection method for optimization of motor imagery based brain–computer interface," *PLoS ONE*, vol. 11, no. 9, Sep. 2016, Art. no. e0162657, doi: 10.1371/journal.pone.0162657.
[28] A. Barachant, "Commande robuste d'un effecteur par une interface cerveau machine EEG asynchrone," Ph.D. thesis, CEA-LETI-Commissariat à L'énergie Atomique Et Aux Énergies Alternatives-Lab. d'Electronique et de Technol. de l'Inf., Univ. de Grenoble, 2012. [Online]. Available: https://theses.hal.science/tel-01196752
[29] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. Hämäläinen, "MEG and EEG data analysis with MNE-Python," *Frontiers Neurosci.*, vol. 7, Dec. 2013, Art. no. 70133, doi: 10.3389/fnins.2013.00267.
[30] G. Pfurtscheller and F. H. L. da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, Nov. 1999, doi: 10.1016/s1388-2457(99)00141-8.
[31] H. Altaheri, G. Muhammad, M. Alsulaiman, S. U. Amin, G. A. Altuwaijri, W. Abdul, M. A. Bencherif, and M. Faisal, "Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review," *Neural Comput. Appl.*, vol. 35, no. 20, pp. 14681–14722, Jul. 2023, doi: 10.1007/s00521-021-06352-5.
[32] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008, doi: 10.1109/msp.2008.4408441.
[33] L. Mesin, U. Ghani, and I. K. Niazi, "Non-linear adapted spatio-temporal filter for single-trial identification of movement-related cortical potential," *Electronics*, vol. 12, no. 5, p. 1246, Mar. 2023, doi: 10.3390/electronics12051246.
[34] L. F. Nicolas-Alonso, R. Corralejo, J. Gómez-Pilar, D. Álvarez, and R. Hornero, "Ensemble learning for classification of motor imagery tasks in multiclass brain computer interfaces," in *Proc. 6th Comput. Sci. Electron. Eng. Conf. (CEEC)*, Sep. 2014, pp. 79–84, doi: 10.1109/CEEC.2014.6958559.
[35] M. Rahimi, A. Zarei, E. Nazerfard, and M. H. Moradi, "Ensemble methods combination for motor imagery tasks in brain computer interface," in *Proc. 23rd Iranian Conf. Biomed. Eng. 1st Int. Iranian Conf. Biomed. Eng. (ICBME)*, Tehran, Iran, Nov. 2016, pp. 336–340, doi: 10.1109/ICBME.2016.7890983.
[36] M. Mohammadpour, M. Ghorbanian, and S. Mozaffari, "Comparison of EEG signal features and ensemble learning methods for motor imagery classification," in *Proc. 8th Int. Conf. Inf. Knowl. Technol. (IKT)*, Hamedan, Iran, Sep. 2016, pp. 288–292, doi: 10.1109/IKT.2016.7777767.
[37] A. Datta, R. Chatterjee, D. K. Sanyal, and D. Guha, "An ensemble classification approach to motor-imagery brain state discrimination problem," in *Proc. Int. Conf. INFOCOM Technol. Unmanned Syst. Trends Future Directions (ICTUS)*, Dubai, United Arab Emirates, Dec. 2017, pp. 322–326, doi: 10.1109/ICTUS.2017.8286026.
[38] A. C. Ramos, R. G. Hernández, M. Vellasco, and P. Vellasco, "Ensemble of classifiers applied to motor imagery task classification for BCI applications," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Anchorage, AK, USA, May 2017, pp. 2995–3002, doi: 10.1109/IJCNN.2017.7966227.
[39] R. Chatterjee, A. Datta, and D. K. Sanyal, "Ensemble learning approach to motor imagery EEG signal classification," in *Machine Learning in Bio-Signal Analysis and Diagnostic Imaging*. New York, NY, USA: Academic, 2019, pp. 183–208, doi: 10.1016/B978-0-12-816086-2.00008-4.
[40] H. Raza, D. Rathee, S.-M. Zhou, H. Cecotti, and G. Prasad, "Covariate shift estimation based adaptive ensemble learning for handling non-stationarity in motor imagery related EEG-based brain–computer interface," *Neurocomputing*, vol. 343, pp. 154–166, May 2019, doi: 10.1016/j.neucom.2018.04.087.
[41] L. Zhang, D. Wen, C. Li, and R. Zhu, "Ensemble classifier based on optimized extreme learning machine for motor imagery classification," *J. Neural Eng.*, vol. 17, no. 2, Mar. 2020, Art. no. 026004, doi: 10.1088/1741-2552/ab7264.
[42] M. N. Cherloo, H. K. Amiri, and M. R. Daliri, "Ensemble regularized common spatio-spectral pattern (ensemble RCSSP) model for motor imagery-based EEG signal classification," *Comput. Biol. Med.*, vol. 135, Aug. 2021, Art. no. 104546, doi: 10.1016/j.compbiomed.2021.104546.
[43] D. K. Nugroho, N. A. Setiawan, and H. A. Nugroho, "Improving multi-class motor imagery EEG signals classification using ensemble learning method," in *Proc. 9th Int. Conf. Inf. Commun. Technol. (ICoICT)*, Yogyakarta, Indonesia, Aug. 2021, pp. 388–393, doi: 10.1109/ICoICT52021.2021.9527426.
[44] M. Rashid, B. S. Bari, M. J. Hasan, M. A. M. Razman, R. M. Musa, A. F. Ab Nasir, and A. P. P. A. Majeed, "The classification of motor imagery response: An accuracy enhancement through the ensemble of random subspace k-NN," *PeerJ Comput. Sci.*, vol. 7, p. e374, Mar. 2021, doi: 10.7717/peerj-cs.374.
[45] L. Zheng, Y. Ma, M. Li, Y. Xiao, W. Feng, and X. Wu, "Time-frequency decomposition-based weighted ensemble learning for motor imagery EEG classification," in *Proc. IEEE Int. Conf. Real-Time Comput. Robot. (RCAR)*, Xining, China, Jul. 2021, pp. 620–625, doi: 10.1109/RCAR52367.2021.9517593.
[46] J. Sun, J. Xie, and H. Zhou, "EEG classification with transformer-based models," in *Proc. IEEE 3rd Global Conf. Life Sci. Technol. (LifeTech)*, Nara, Japan, Mar. 2021, pp. 92–93, doi: 10.1109/LifeTech52111.2021.9391844.
[47] S. Kamhi, S. Zhang, M. A. Amou, M. Mouhafid, I. Javaid, I. S. Ahmad, I. A. El Kader, and U. Kulsum, "Multi-classification of motor imagery EEG signals using Bayesian optimization-based average ensemble approach," *Appl. Sci.*, vol. 12, no. 12, p. 5807, Jun. 2022, doi: 10.3390/app12125807.
[48] Q. Wang, Y. Huo, Z. Xu, W. Zhang, Y. Shang, and H. Xu, "Effects of backrest and seat-pan inclination of tractor seat on biomechanical characteristics of lumbar, abdomen, leg and spine," *Comput. Methods Biomech. Biomed. Eng.*, vol. 26, no. 3, pp. 291–304, 2023, doi: 10.1080/10255842.2022.2062229. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/10255842.2023.2284091
[49] D.-H. Shin, Y.-H. Son, J.-M. Kim, H.-J. Ahn, J.-H. Seo, C.-H. Ji, J.-W. Han, B.-J. Lee, D.-O. Won, and T.-E. Kam, "MARS: Multiagent reinforcement learning for spatial–spectral and temporal feature selection in EEG-based BCI," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 54, no. 5, pp. 3084–3096, May 2024, doi: 10.1109/TSMC.2024.3355101.
[50] A. K. Singh and S. Krishnan, "Trends in EEG signal feature extraction applications," *Frontiers Artif. Intell.*, vol. 5, Jan. 2023, Art. no. 1072801, doi: 10.3389/frai.2022.1072801.
[51] V. Arora, L. Behera, T. K. Reddy, and A. P. Yadav, "HJB equation based learning scheme for neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Anchorage, AK, USA, May 2017, pp. 2298–2305, doi: 10.1109/IJCNN.2017.7966134.

HOSSEIN AHMADI received the B.S. degree in electronics engineering from the University of Kurdistan, Sanandaj, Iran, in 2010, and the M.S. degree in telecommunications engineering from the Amirkabir University of Technology, Tehran, Iran, in 2016. He is currently pursuing the Ph.D. degree with the Politecnico di Torino, Italy, specializing in electrical, electronics, and communications engineering. His research interests include brain-to-brain communication, semantic communication, and signal processing.

LUCA MESIN received the bachelor's degree in electronics engineering, in 1999, and the Ph.D. degree in applied mathematics from the Politecnico di Torino, Italy, in 2003. He is currently an Associate Professor of biomedical engineering and the Supervisor of the Mathematical Biology and Physiology Group, Department of Electronics and Telecommunications, Politecnico di Torino. His research interests include biomedical image and signal processing and mathematical modeling.

Open Access funding provided by 'Politecnico di Torino' within the CRUI CARE Agreement