Journal of Theoretical and Applied Information Technology
31st December 2025. Vol.103. No.24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CONSENSUS LEARNING WITH PATHOPHYSIOLOGY-INFORMED SPECTRAL FEATURES FOR EEG-BASED STROKE AND TBI DIAGNOSIS

GAO JIALU[1], NORWATI MUSTAPHA[1], NORIDAYU BINTI MANSHOR[1], RAZALI BIN YAAKOB[1]

[1]Faculty Of Computer Science And Information Technology, Universiti Putra Malaysia, Malaysia

E-mail: [1]gs67746@student.upm.edu.my

ABSTRACT

Traumatic brain injury (TBI) and stroke are distinct yet critical neurological conditions requiring timely and accurate diagnosis. While EEG-based binary classification models often exceed 90% accuracy, clinically relevant three-class classification (TBI vs. stroke vs. normal) remains a significant challenge, with current state-of-the-art methods achieving only 69% accuracy and 0.84 ROC-AUC. This study introduces a novel consensus learning framework that incorporates pathophysiology-informed frequency band decomposition to improve classification performance. Three specialized spectral configurations are employed: a TBI-optimized delta band (0.5–4 Hz), a stroke-sensitive alpha band (8–13 Hz), and a cross-pathology band covering conventional clinical frequencies. Identical TMN-CNN architectures are trained on each spectral view, and an adaptive consensus mechanism integrates predictions via class-specific weighting and confidence-based voting. Evaluated on 9,276 EEG segments from the Temple University Hospital EEG Corpus, the proposed method achieves 76.57% balanced accuracy and 0.911 ROC-AUC, outperforming benchmarks by +0.071 ROC-AUC. Class-specific analysis confirms balanced performance across categories, with a +3.22% gain over the best individual model. These results demonstrate the effectiveness of pathophysiology-guided, multi-view consensus learning for robust EEG-based neurological classification.

Keywords: *EEG Classification, Traumatic Brain Injury, Stroke Detection, Consensus Learning, Multi-Modal Signal Integration)*

1. INTRODUCTION

Traumatic brain injury (TBI) and stroke are among the leading causes of neurological disability and death worldwide [1,2]. Timely and accurate diagnosis is critical—especially for stroke, where intervention within a 3–4.5-hour window can significantly improve outcomes [3]. However, conventional imaging techniques like CT and MRI are often limited by high costs, low accessibility in emergency and rural settings, and delayed results.

Electroencephalography (EEG) has emerged as a promising alternative due to its portability, affordability, and high temporal resolution [4]. While most EEG-based studies focus on binary classification (e.g., TBI vs. normal or stroke vs. normal), achieving accuracies above 90% [5], the clinically essential three-class classification (TBI vs. stroke vs. normal) remains more challenging, with current models achieving only moderate performance. For instance, Caiola et al. [6] achieving an ROC-AUC of 0.84 in distinguishing between TBI, stroke, and normal patterns using neural networks with EEG data from a nonspecific population demonstrating the benchmark performance for this complex classification task. However, their single-model approach struggles with the variability and complexity of EEG patterns across different pathologies.

Critically, existing methods overlook condition-specific spectral signatures. TBI typically involves increased delta (0.5–4 Hz) and decreased alpha activity, reflecting slow-wave pathology [6], while stroke often presents alpha band disruptions and hemispheric asymmetries [8,9]. Most current approaches apply uniform spectral analyses, failing to exploit these pathophysiological distinctions [7–9]. Additionally, ensemble methods in EEG classification generally combine different architectures on identical features, missing the opportunity to leverage complementary spectral views [10].

To address these limitations, this study introduces a **pathophysiology-informed consensus learning framework**. Unlike traditional ensembles, we employ identical TMN-CNN architectures trained on three complementary spectral views derived from neurophysiological insights:

1. **TBI-Optimized Band** (0.5–2 Hz for deep injuries, 2–4 Hz for superficial damage);
2. **Stroke-Sensitive Band** (8–10 Hz and 10–13 Hz for hemispheric asymmetry detection);
3. **Cross-Pathology Band** (broad-spectrum features using conventional clinical bands).

A **weighted consensus voting mechanism** then integrates the predictions, accounting for class-specific expertise and model confidence. This multi-view, pathophysiology-driven approach enhances model specialization and improves diagnostic accuracy for complex neurological classification tasks.

2. LITERATURE REVIEW

2.1 Current Approaches And Limitations

Recent advancements in machine learning have shown strong performance in EEG-based neurological classification, particularly in binary settings. For TBI detection, Vivaldi et al. achieved 94% accuracy using data-driven methods with SVM and K-nearest neighbors on EEG signals [11]. In stroke classification, Tong et al. reported 96.76% precision and 96.69% recall using their ApFu-TPELGBM model, claiming it as the most accurate EEG-based stroke classifier to date [12]. These results highlight the high potential of EEG-based diagnostics in binary classification scenarios.

Advances in deep learning have significantly enhanced EEG-based stroke classification. Lin et al. introduced AM-EEGNet, a multi-input model based on EEGNet that integrates power spectral density and functional connectivity, achieving 97.22% accuracy, 0.94 sensitivity, and 1.00 specificity [13]. Traditional feature-based methods also show promising by a hybrid model combining RNN and CNN architectures with harmony search-based feature selection reported 99.991% accuracy [14]. While effective, these approaches focus primarily on binary classification tasks.

However, performance drops significantly when moving from binary to clinically relevant multi-class classification. For example, Vivaldi et al.'s accuracy declined from 94% in binary TBI detection to 71% in a three-class setting (TBI vs. stroke vs. normal) during independent validation [11]. This gap underscores a key limitation in current methods, echoed by Yuan et al., whose multi-view deep learning framework in combining unsupervised EEG reconstruction with supervised spectrogram-based detection was still struggles towards multi-class complexity [15]. Similarly, Caiola et al.'s benchmark ROC-AUC of 0.84 in three-class classification [6], while representing the state of the art, highlights the performance disparity compared to binary models.

Current EEG diagnostic systems face a critical limitation: while achieving high accuracy in binary classification tasks, they struggle with three-class differentiation due to their reliance on uniform feature extraction and single-model architectures. This performance disparity highlights the urgent need for frameworks capable of simultaneously exploiting multiple pathology-specific spectral signatures, which our consensus learning approach addresses through complementary frequency band representations and model-level diversity.

2.2 Consensus Learning

Ensemble learning has been used to address the limitations of single-model EEG classifiers, but current approaches face a key conceptual constraint. Most ensembles rely on model diversity in which either through heterogeneous architectures or multiple instances of the same model. This strategy has proven effective in related domains, such as deep learning ensembles for ECG-based heart disease detection [16]. In EEG analysis, hybrid models combining CNNs and RNNs have also shown promise, leveraging their distinct learning biases to capture complementary signal features [17].

A key limitation of current ensemble frameworks lies in their reliance on a single, uniform feature representation, despite model diversity. This single-view approach constrains the ensemble's effectiveness, as all models regardless of architecture that make decisions based on the same perspective of the EEG signal. The issue is not model complexity, but the lack of feature diversity. Although **Consensus Learning** promotes combining insights from multiple distinct views for more robust decisions [18], this data-centric strategy remains underexplored in EEG

analysis. As a result, there is a critical gap in systematically developing ensembles of specialized models trained on deliberately crafted, complementary feature sets.

2.3 Band Frequency

The limitations of current feature extraction methods are evident when considering the distinct neurophysiological signatures of TBI and stroke. Most approaches rely on generic spectral analysis, which overlooks pathology-specific biomarkers. In TBI, increased delta power (0.5–4 Hz) is a key marker, yet it's often treated as a single, undifferentiated band. This ignores evidence that lower delta frequencies may indicate deeper white matter damage, while higher frequencies suggest superficial cortical injury [19]. Without stratifying the delta band, critical diagnostic information is lost.

Similarly, in ischemic stroke, hemispheric asymmetry in the alpha band (8–13 Hz) is a primary feature. However, this band includes functionally distinct sub-bands—such as lower and upper alpha—that are differentially affected by stroke [20]. Averaging across the full alpha range can obscure subtle, yet clinically relevant, asymmetries. Overall, current methods fail to align spectral analysis with the underlying pathophysiology of each condition. This highlights a key research gap: the need for multi-view analytical frameworks with pathology-specific spectral feature engineering to better capture distinct neurological biomarkers.

Current EEG analysis methods face limitations not only in pathology-specific modeling but also in addressing individual variability in frequency bands. Neurofeedback research shows that applying standardized bands without accounting for individual EEG profiles can activate irrelevant neural activity [21]. For instance, theta and beta ranges can be more accurately defined using each person's individual alpha peak frequency, and studies suggest that personalized bands improve analytical precision over fixed ranges [21]. Ignoring both inter-individual variability and pathology-specific features thus compromises the diagnostic accuracy of EEG-based approaches.

Emerging evidence indicates that neurological conditions often involve complex interactions across multiple frequency bands rather than isolated spectral changes. For example, in chronic pain, increased frontocentral beta power alongside reduced theta activity correlates with disease severity [22]. Similarly, the theta/beta ratio has proven more diagnostically informative than individual bands, showing inverse relationships with cognitive control across various neurological and psychiatric disorders [23]. These findings highlight the need for analytical frameworks that capture multi-frequency interactions while remaining sensitive to pathology-specific patterns. Developing cross-pathology approaches that detect broad-spectrum abnormalities without losing specificity is a key unmet need in EEG-based diagnostics.

3. METHODOLOGY

3.1 Dataset Description

This study used a subset of the Temple University Hospital EEG (TUH EEG) Corpus to perform three-class classification of traumatic brain injury (TBI), stroke, and normal controls. Diagnostic labels were obtained from Caiola et al. [6], who provided clinically validated annotations based on comprehensive medical record reviews.

The full TUH EEG Corpus (v2.0.1) contains 16,986 sessions from 10,874 subjects, recorded at Temple University Hospital between 2002 and 2017 [24]. For this study, only recordings with confirmed TBI, stroke, or normal labels from Caiola et al. were included. After preprocessing and segmentation, the final dataset comprised 9,276 three-minute EEG segments from 1,390 subjects across 1,571 sessions, with balanced representation across classes (see Table 1).

EEG recordings followed standard clinical protocols using the international 10–20 electrode placement system. Sampling rates ranged from 250 to 512 Hz, reflecting typical clinical EEG acquisition settings. The TUH EEG Corpus utilized in this study is publicly available and contains de-identified patient information released with appropriate institutional approvals. All data were anonymized prior to public release, with diagnostic labels obtained through validated clinical record reviews by Caiola et al. This research involves secondary analysis of publicly accessible, anonymized data and complies with established ethical guidelines for human subjects research.

Table 1 Dataset composition and demographics

Category	Subjects	Recordings	Age (years)	Sex (M/F)	Segments
Normal	498	527	42.2 ± 13.4	336/181	3091
TBI	471	572	43.2 ± 13.7	354/150	3351
Stroke	421	472	45.6 ± 12.2	311/152	2834
Total	**1390**	**1571**	**43.6 ± 13.1**	**1001/483**	**9276**

3.2 Data Preprocessing

EEG recordings were preprocessed using a standardized pipeline to isolate brain-related activity through systematic artifact removal. Data were standardized to a 19-channel montage (10–20 system), with multi-file sessions concatenated into continuous recordings. The first 60 seconds were discarded to remove stabilization artifacts, and recordings were segmented into non-overlapping 3-minute epochs.

Signals were bandpass filtered between 1–100 Hz using finite impulse response (FIR) filters and re-referenced to the common average across all channels. Each 3-minute segment was decomposed using Independent Component Analysis (ICA) with the extended Infomax algorithm to extract independent sources. Components were classified using ICLabel [25], which assigns probabilities across seven categories (e.g., brain, muscle, eye, cardiac, noise). The "Only Brain" method was applied to retain only components primarily classified as brain activity, ensuring reconstructed signals were free of non-neural artifacts.

3.3 Feature Extraction

This study introduces a novel parallel frequency band configuration strategy to enhance feature extraction for multi-class neurological diagnosis. Three specialized configurations which are **TBI-Optimized Band**, **Stroke-Sensitive Band**, and **Cross-Pathology Band** were developed in parallel, each generating independent topographic map networks (TMNs) to capture complementary brain activity patterns for consensus classification.

- **TBI-Optimized Band** focuses on delta stratification with seven bands: Delta Deep (0.5–2 Hz) and Delta Mild (2–4 Hz) for deep vs. superficial injury, Theta Low (4–6 Hz) and Theta High (6–8 Hz) for TBI-specific patterns, Alpha Low (8–10 Hz) and Alpha High (10–13 Hz) for baseline comparison, and Beta-Gamma Combined (13–50 Hz) for high-frequency activity with limited static discriminative value.
- **Stroke-Sensitive Band** emphasizes alpha subdivision and hemispheric asymmetry, including Delta Slow (0.5–2 Hz) and Delta Fast (2–4 Hz), Theta Pathological (4–7 Hz), Alpha Low (8–10 Hz) and Alpha High (10–13 Hz) for asymmetry detection, Beta Disrupted (13–25 Hz), and Gamma (30–50 Hz) for compensatory activity.
- **Cross-Pathology Band** applies standard clinical frequency divisions—Delta (1–4 Hz), Theta (4–8 Hz), Alpha Low (8–10 Hz), Alpha High (10–13 Hz), Beta Low (13–20 Hz), Beta High (20–30 Hz), and Gamma (30–50 Hz)—to support baseline comparisons and validate pathology-specific configurations.

A detailed comparison of these frequency band configurations is provided in **Table 2**.

Table 2 Detailed frequency band configurations comparison

Band	TBI-Optimized Band	Stroke-Sensitive Band	Cross-Pathology Band	Clinical Rationale
1	Delta Deep (0.5-2 Hz)	Delta Slow (0.5-2 Hz)	Delta (1-4 Hz)	Deep injury vs. baseline pathological activity
2	Delta Mild (2-4 Hz)	Delta Fast (2-4 Hz)	Theta (4-8 Hz)	Superficial damage vs. mild stroke abnormalities
3	Theta Low (4-6 Hz)	Theta Pathological (4-7 Hz)	Alpha Low (8-10 Hz)	TBI-specific theta patterns
4	Theta High (6-8 Hz)	Alpha Low (8-10 Hz)	Alpha High (10-13 Hz)	Normal theta vs. hemispheric asymmetry
5	Alpha Low (8-10 Hz)	Alpha High (10-13 Hz)	Beta Low (13-20 Hz)	Baseline comparison vs. critical stroke detection
6	Alpha High (10-13 Hz)	Beta Disrupted (13-25 Hz)	Beta High (20-30 Hz)	Baseline comparison vs. stroke coherence patterns
7	Beta-Gamma Combined (13-50 Hz)	Gamma (30-50 Hz)	Gamma (30-50 Hz)	Reduced static value vs. compensatory activity

Power spectral density (PSD) was computed for each frequency configuration using Welch's method with a 2048-sample window. Band-specific power was extracted by integrating over each defined frequency range across all 19 EEG channels. Electrode positions were projected from 3D space onto a 2D 134×134 pixel grid using spherical projection with margin spacing.

Spatial interpolation was performed using cubic radial basis function (RBF) interpolation with smoothing parameter $\sigma = 0.1$, generating continuous topographic maps from discrete electrode data. The interpolation function is defined as equation (1):

$$f(x,y) = \Sigma_i w_i \varphi(||r - r_i||) \quad (1)$$

where $\varphi(r)$ represents the cubic basis function $\varphi(r) = r^3$, w_i are the interpolation weights determined by solving the linear system, and r_i are the electrode positions. Circular masking approximating head geometry eliminated interpolated values outside physiologically relevant scalp regions, ensuring biologically meaningful topographic representations. Each frequency configuration produced independent 134×134×7 TMN tensors, preserving both spectral and spatial information while maintaining computational efficiency for deep learning applications. These TMN representations serve as input features for the subsequent consensus CNN architecture, enabling the model to leverage complementary neurological pathology signatures from each specialized frequency configuration. The complete proposed consensus learning framework is illustrated in Figure 1.

Journal of Theoretical and Applied Information Technology
31st December 2025. Vol.103. No.24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

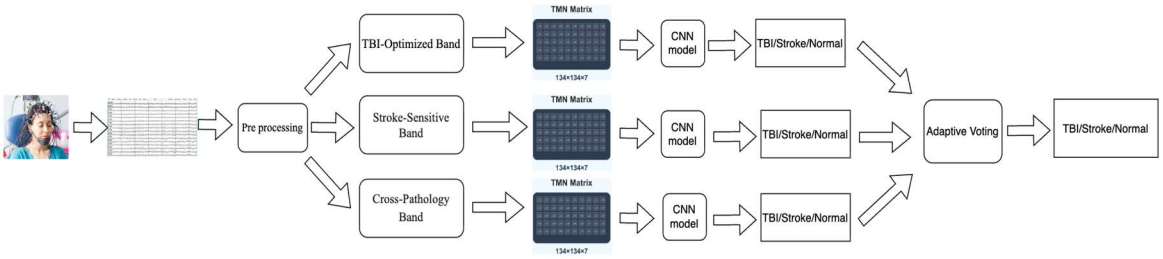

Figure 1: The proposed consensus learning framework

3.4 Model Architecture

The classification system uses an enhanced Topographic Map Network (TMN) architecture optimized for multi-class EEG pathology detection. It processes 134×134×7 topographic inputs, where each channel represents spatial power from a specific frequency band, as defined in Section 3.3.

The architecture consists of three main components:

- **Feature Extraction Module**: Begins with a 1×1 convolution (7→64 channels), followed by batch normalization and ReLU to model inter-frequency relationships. The network uses channel-wise normalization during preprocessing and per-channel batch normalization after each convolution. It employs alternating 1×1 and 3×3 convolutions with increasing depth (up to 128 channels) and spatial downsampling via max pooling (3×3, 5×5, 5×5, stride 2).
- **Adaptive Pooling**: Reduces spatial dimensions to a fixed 8×8, ensuring consistent feature representation regardless of input variability.
- **Classification Head**: Consists of three fully connected layers (8192→256→100→3) with ReLU activations and dropout (rates 0.3 and 0.21) to prevent overfitting. The final output layer produces logits for the three classes: Normal, TBI, and Stroke.

The full TMN-CNN architecture is illustrated in **Figure 2**.

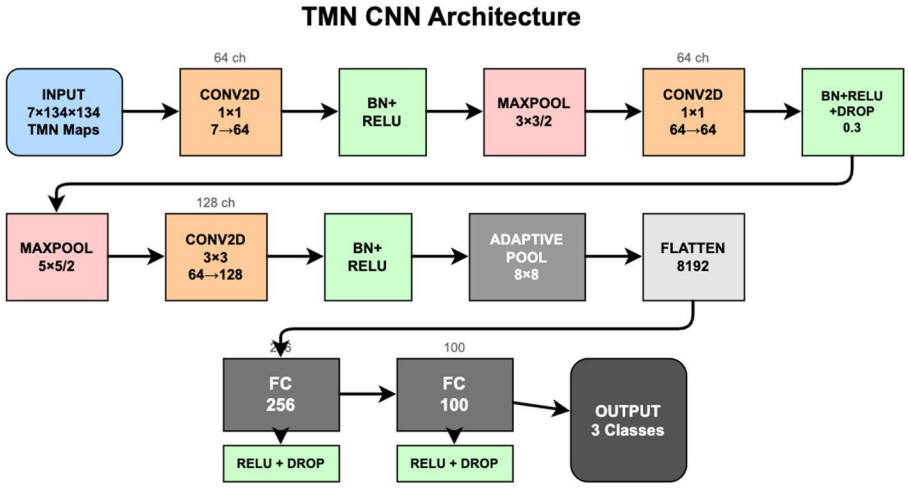

Figure 2 TMN CNN architecture

The model applies batch normalization across all convolutional layers to stabilize training and enhance convergence. Weights are initialized using Kaiming normal for convolutional layers and a normal distribution for linear layers, resulting in approximately 2.1 million trainable parameters. Training uses Adam optimizer with a learning rate of 5×10^{-5}, weight decay of 1×10^{-5}, and mixed precision for efficiency. The architecture balances computational efficiency with preservation of spatial features essential for neurological pathology detection. Due to slight class imbalance (Normal: 3,091; TBI: 3,351; Stroke: 2,834 segments), balanced accuracy is used as the primary evaluation metric to ensure fair performance assessment across classes.

3.5 Consensus Learning Strategy

The consensus system combines three identical TMN models, each trained on distinct frequency band configurations to capture complementary neurological pathology features. This multi-model approach leverages the specialized strengths of each configuration while ensuring robustness via adaptive voting.

- The TBI-Optimized Model focuses on delta stratification (0.5-2 Hz and 2-4 Hz) to detect TBI-specific slow waves and connectivity disruptions.
- The Stroke-Sensitive Model targets stroke-related features using alpha sub-bands (8-10 Hz, 10-13 Hz) and beta disruption (13-25 Hz) to capture cerebrovascular markers and hemispheric asymmetries.
- The Cross-Pathology Model employs standard clinical frequency bands for broad neurophysiological assessment.

The voting mechanism dynamically weights predictions based on model expertise and confidence. Class-Specific Expertise Weighting assigns tailored weights for Normal ([0.45, 0.30, 0.25]), Stroke ([0.55, 0.25, 0.20]), and TBI ([0.30, 0.35, 0.35]) reflecting each model's domain proficiency. Confidence-Based Adjustment scales weights according to softmax scores, boosting predictions with confidence >0.8 and down-weighting those <0.6. Consensus-Driven Integration applies majority voting with confidence weighting when agreement exceeds 67%; otherwise, it uses weighted probability averaging informed by expertise and historical accuracy.

This hierarchical consensus framework harnesses frequency-band specialization and adaptive weighting to enhance classification accuracy across normal, stroke, and TBI conditions. The consensus mechanism operates through a sequential decision pipeline: first, each model generates class probabilities via softmax activation; second, expertise weights are applied based on the predicted class label; third, confidence adjustment modulates these weights according to prediction certainty thresholds; finally, the weighted ensemble decision is determined either through majority voting (when model agreement ≥67%) or weighted probability summation. This implementation ensures that models with demonstrated strength in specific pathology detection contribute proportionally more to the final classification while maintaining system robustness through distributed decision-making. The voting mechanism itself adds negligible computational cost compared to model inference, making the framework suitable for clinical deployment scenarios.

4. RESULTS

4.1 Overall Performance Comparison

The consensus learning approach demonstrates substantial improvements in three-class neurological classification performance, addressing the critical gap identified in current EEG-based diagnostic methods. The proposed consensus learning strategy achieved a balanced accuracy of 76.57% with an average ROC-AUC of 0.91, representing significant advancement over existing state-of-the-art approaches for multi-class TBI, stroke, and normal classification tasks.

The comprehensive performance comparison presented in Table 3 reveals the progressive improvements achieved through the multi-view frequency band decomposition strategy. Individual specialized models showed encouraging results, with the Cross-Pathology configuration achieving 71.52% accuracy and 0.886 average ROC-AUC, while both TBI-focused configurations (Stroke-Sensitive and TBI-Optimized) demonstrated similar performance levels of 73.35% and 73.14% accuracy respectively, with average ROC-AUC values of 0.889 and 0.892. However, the true strength of the proposed approach emerges through the Adaptive Voting consensus mechanism, which integrates these complementary models to achieve 76.38% accuracy with 0.911 average ROC-AUC.

Journal of Theoretical and Applied Information Technology
31st December 2025. Vol.103. No.24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3 Comprehensive Performance Comparison with Benchmarks

Model	Overall balance Accuracy	F1-Macro	ROC-AUC (Normal)	ROC-AUC (Stroke)	ROC-AUC (TBI)	ROC-AUC (avg)
Caiola et al. [6] best model	-	-	0.86	0.84	0.81	0.84
Caiola et al. [6] TMN model	-	-	0.80	0.75	0.71	0.76
Cross-Pathology	71.52%	0.715	0.900	0.906	0.851	0.886
Stroke-Sensitive	73.35%	0.733	0.901	0.909	0.857	0.889
TBI-Optimized	73.14%	0.730	0.900	0.906	0.869	0.892
Adaptive Voting	**76.57%**	**0.763**	**0.922**	**0.925**	**0.885**	**0.911**
Improvement	-	-	**+0.082**	**+0.085**	**+0.045**	**+0.071**

*Balanced accuracy is reported to account for class imbalance

Compared to the benchmark by Caiola et al. [6], the consensus approach achieves significant improvements across all metrics. While the benchmark reported an ROC-AUC of 0.84 for three-class classification, the adaptive voting method attains an average ROC-AUC of 0.911, a +0.071 increase. It also outperforms the best individual models with a +3.03% accuracy gain, +0.030 improvement in F1-macro, and enhanced class-specific ROC-AUC across all categories.

Class-wise ROC-AUC improvements over the best models are +0.021 for Normal, +0.016 for Stroke, and +0.016 for TBI, demonstrating the effectiveness of the pathophysiology-informed frequency band decomposition in overcoming multi-class classification challenges. The F1-macro score of 0.763 further confirms balanced performance across classes, addressing typical clinical dataset imbalances.

Such balanced and robust classification is critical for clinical applications, ensuring high sensitivity and specificity across all pathological categories for reliable patient care.

4.2 Consensus Learning Strategy Analysis

The consensus learning approach excels by integrating complementary insights from three specialized models, each trained on pathophysiology-informed frequency bands. Unlike traditional ensembles that combine diverse architectures on the same features, this method uses a consistent TMN-CNN architecture while leveraging distinct spectral representations tailored to specific neurological pathologies. This multi-view strategy allows each model to develop focused expertise on condition-specific biomarkers while ensuring seamless integration.

The three models offer synergistic diagnostic strengths: the TBI-Optimized model targets pathological slow waves via delta stratification (0.5-2 Hz for deep, 2-4 Hz for superficial injury), the Stroke-Sensitive model detects hemispheric asymmetries through alpha subdivision (8-10 Hz, 10-13 Hz), and the Cross-Pathology model captures broad-spectrum abnormalities using standard clinical bands. This diversification ensures unique, complementary neurophysiological insights that collectively encompass the full diagnostic spectrum of the three conditions.

Table 4: Consensus Strategy Comparison

Consensus Method	Accuracy	F1-Macro	Improvement over Best Single
Simple Average	75.73%	0.757	+2.38%
F1-Weighted Consensus	75.84%	0.758	+2.49%
Enhanced Adaptive	75.94%	0.759	+2.59%
Adaptive Consensus	**76.57%**	**0.763**	**+3.22%**

Table 4 clearly demonstrates the superiority of the Adaptive Consensus strategy over alternative voting methods, with consistent improvements across all metrics. Simple averaging yields a baseline accuracy of 75.73% but ignores individual model expertise and confidence. F1-weighted consensus slightly improves to 75.84% by using static historical weights, while the Enhanced Adaptive method, incorporating confidence-based adjustments, reaches 75.94%.

The full Adaptive Consensus mechanism outperforms all alternatives, achieving 76.57% balanced accuracy and a 0.763 F1-macro score—an improvement of +3.22% over the best single model. This gain stems from dynamically combining three elements: class-specific expertise weighting based on each model's strengths, confidence-based modulation of predictions via softmax scores, and consensus-driven integration that applies majority voting with confidence weighting above 67% agreement and weighted probability averaging otherwise.

This adaptive approach is particularly effective for the heterogeneous EEG patterns in neurological pathologies, where rigid voting fails to capture subtle diagnostic signatures distinguishing TBI, stroke, and normal states. By adjusting decision weights dynamically, it optimally integrates complementary pathophysiological insights while maintaining robustness against individual model uncertainty.

Table 5: Detailed Classification Metrics

Class	Precision	Recall	F1-Score	Class balance Accuracy
Normal	0.795	0.727	0.760	81.63%
Stroke	0.753	0.846	0.796	77.78%
TBI	0.745	0.724	0.734	70.31%
Macro Avg	**0.764**	**0.766**	**0.763**	**76.57%**

*Balanced accuracy reported to ensure fair evaluation across classes with different sample sizes

Normal classification achieves the highest precision (0.795) but lower recall (0.727), yielding an F1-score of 0.760 and class accuracy of 81.63%. This indicates a conservative bias, minimizing false positives to avoid unnecessary clinical interventions. High precision ensures reliable normal predictions, while lower recall suggests some normal cases are misclassified as pathological.

Stroke detection shows the most balanced profile, with moderate precision (0.753) and the highest recall (0.846), resulting in the best F1-score (0.796) and class accuracy (77.78%). The high recall is clinically critical to minimize missed stroke diagnoses, reflecting the Stroke-Sensitive model's effective alpha subdivision strategy for detecting hemispheric asymmetries typical of cerebrovascular pathology (see Figures 4a and 4b).

TBI classification is the most challenging, with the lowest precision (0.745), recall (0.724), F1-score (0.734), and class accuracy (70.31%). This likely reflects TBI's heterogeneous pathology, ranging from focal cortical lesions to diffuse axonal injury, complicating EEG-based detection and requiring more advanced analytical methods (refer to Figures 3a and 3b).

Journal of Theoretical and Applied Information Technology
31st December 2025. Vol.103. No.24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

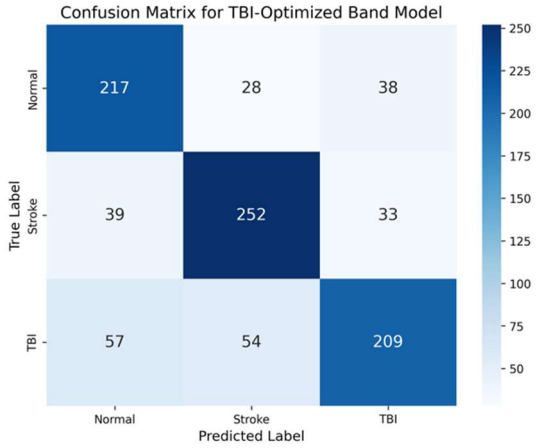

Figure 3a: TBI-Optimized model confusion matrix

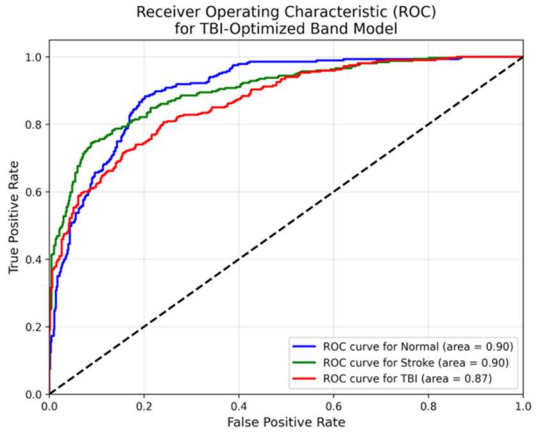

Figure 3b: TBI-Optimized model ROC curves

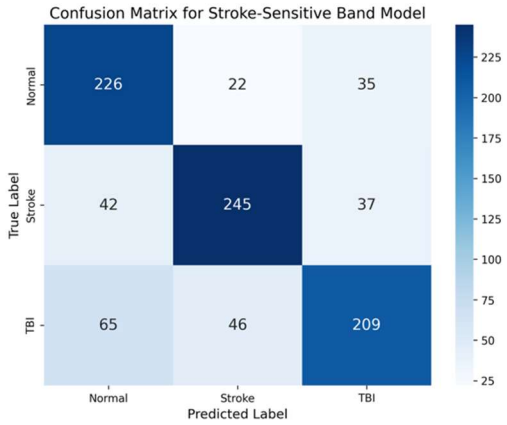

Figure 4a: Stroke-Sensitive model confusion matrix

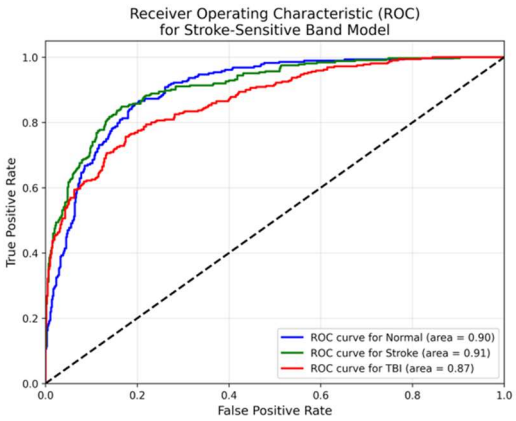

Figure 4b: Stroke-Sensitive model ROC curves

Individual model performance is illustrated through confusion matrices and ROC curves: Figures 3a and 3b show the TBI-Optimized model, Figures 4a and 4b the Stroke-Sensitive model, and Figures 5a and 5b the Cross-Pathology model. These visualizations highlight each model's specialized diagnostic strengths and limitations, reflecting their unique contributions to the consensus framework. The Cross-Pathology model provides baseline performance across standard clinical frequency bands (Figures 5a and 5b).

Figures 6a and 6b present the Adaptive Voting consensus results, demonstrating superior discrimination and consistently high AUC values across all three classes, validating the effectiveness of the pathophysiology-informed consensus learning approach.

Class-specific analysis emphasizes the clinical relevance of this method, where specialized frequency bands enhance detection of distinct conditions, and the consensus mechanism optimally balances precision and recall based on the clinical importance of each pathology. This tailored performance supports the system's utility as an adaptable diagnostic tool responsive to specific neurological conditions.

Journal of Theoretical and Applied Information Technology
31st December 2025. Vol.103. No.24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

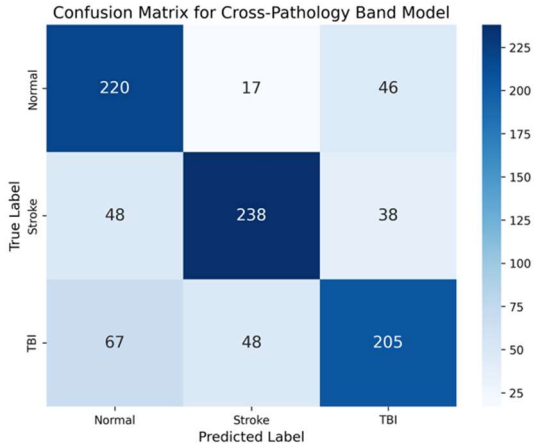

Figure 5a: Cross-pathology model confusion matrix

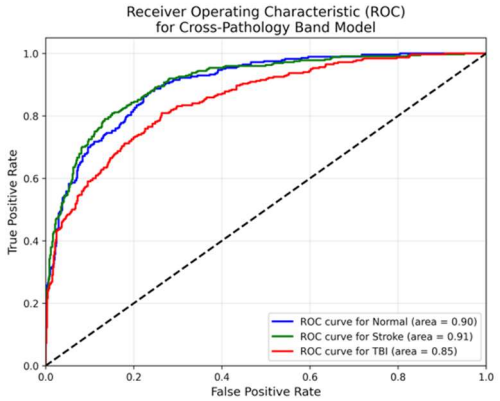

Figure 6b: Cross-pathology model ROC curves

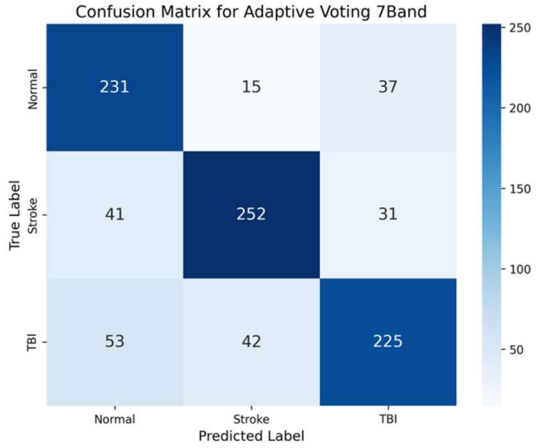

Figure 6a: Adaptive Voting confusion matrix

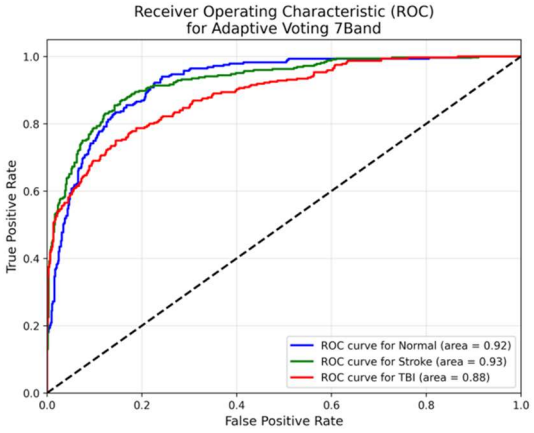

Figure 6b: Adaptive Voting ROC curves

5. CONCLUSION

This study introduces a novel consensus learning framework for EEG-based three-class neurological classification, effectively bridging the performance gap between binary and multi-class diagnostics. The pathophysiology-informed frequency band decomposition strategy achieves 76.57% accuracy and an average ROC-AUC of 0.911, surpassing the previous benchmark ROC-AUC of 0.84 by Caiola et al.

The key innovation is the use of three specialized frequency band configurations which are TBI-Optimized Band (delta stratification), Stroke-Sensitive Band (alpha subdivision), and Cross-Pathology Band (broad-spectrum) to exploit condition-specific spectral signatures. Unlike traditional ensembles combining diverse architectures on the same features, this approach employs identical TMN-CNN models trained on complementary spectral inputs, enabling specialized expertise with architectural consistency.

The adaptive consensus voting mechanism enhances performance by +3.22% over the best single model through dynamic weighting based on class-specific expertise and prediction confidence. Class-wise results show high precision for Normal (0.795), optimal recall for Stroke (0.846), and improved robustness for TBI despite its heterogeneous pathology.

Clinically, this approach overcomes limitations in current EEG diagnostics by integrating neurophysiology-driven multi-view feature engineering with scalable consensus learning. It

supports condition-specific analysis without added architectural complexity or excessive computation.

Future work includes expanding the framework to other neurological disorders, incorporating temporal EEG dynamics, validating across diverse clinical settings, and developing real-time and personalized frequency band implementations to enhance diagnostic precision and utility.

This pathophysiology-informed consensus learning framework lays a foundation for advancing EEG-based neurological diagnosis towards more accurate, reliable, and clinically actionable tools, facilitating timely intervention and better patient outcomes in critical neurological conditions.

Beyond immediate clinical applications, this work holds broader implications for healthcare accessibility and equity. The portability and affordability of EEG compared to CT and MRI make this framework particularly valuable for resource-limited settings, rural healthcare facilities, and emergency medical services where advanced neuroimaging may be unavailable or delayed. By enabling rapid, accurate differential diagnosis using widely accessible technology, the proposed system has potential to reduce disparities in neurological care access and improve outcomes in underserved populations.

The methodological contributions extend beyond stroke and TBI detection. The principle of pathophysiology-informed feature decomposition combined with consensus learning represents a generalizable paradigm applicable to other medical diagnostic challenges involving heterogeneous conditions with distinct physiological signatures. Potential applications include cardiology (ECG-based multi-class arrhythmia detection), pulmonology (respiratory signal analysis for various pulmonary conditions), and psychiatry (EEG-based classification of different mental health disorders).

Moving forward, translating this research into clinical practice requires collaborative efforts across multiple stakeholders. We call upon medical device manufacturers to explore integration of such intelligent diagnostic algorithms into portable EEG systems. Healthcare institutions are encouraged to conduct prospective validation studies assessing real-world diagnostic accuracy and clinical workflow integration. Regulatory bodies should consider pathways for evaluation and approval of AI-assisted diagnostic tools that can enhance but not replace clinical judgment. Finally, the research community must prioritize standardization of EEG acquisition protocols and establishment of large-scale, multi-center datasets to facilitate robust model development and validation.

6. DISCUSSION

6.1 Interpretation of Results

The proposed consensus learning framework demonstrates substantial improvements over existing approaches, achieving 76.57% balanced accuracy and 0.911 ROC-AUC compared to the benchmark ROC-AUC of 0.84 [6]. This +0.071 improvement in ROC-AUC, while seemingly modest numerically, represents clinically significant progress in a challenging three-class classification task where binary methods typically exceed 90% accuracy but struggle with multi-class scenarios.

The high recall for stroke detection (0.846) is particularly noteworthy from a clinical perspective, as minimizing false negatives is critical given the time-sensitive nature of stroke intervention. The stroke-sensitive model's alpha subdivision strategy (8-10 Hz, 10-13 Hz) effectively captures hemispheric asymmetries characteristic of cerebrovascular pathology, contributing to this strong performance. Conversely, TBI classification remains the most challenging (70.31% class accuracy), reflecting the heterogeneous nature of traumatic brain injuries ranging from focal cortical lesions to diffuse axonal injury.

6.2 Limitations

Several limitations should be acknowledged. First, the dataset originates from a single institution (Temple University Hospital), which may limit generalizability across different clinical settings, patient populations, and EEG acquisition protocols. Multi-center validation studies are needed to establish robustness across diverse clinical environments.

Second, the frequency band configurations were predefined based on existing neurophysiological knowledge rather than learned adaptively from data. While this pathophysiology-informed approach demonstrates effectiveness, individual variations in

brain rhythms—such as personalized alpha peak frequencies—are not currently accommodated. Future implementations incorporating adaptive, patient-specific frequency band tuning may further improve diagnostic accuracy.

Third, the current framework requires 3-minute EEG segments for analysis. While this duration is suitable for comprehensive assessment in clinical monitoring settings, shorter analysis windows may be preferable for emergency triage scenarios. Investigating optimal time window configurations for different clinical contexts represents an important avenue for future work.

Finally, the consensus mechanism's class-specific weights were empirically determined based on validation performance. More systematic approaches, such as automated hyperparameter optimization or meta-learning strategies, could potentially enhance the adaptability of the voting mechanism.

6.3 Future Directions

Several promising directions emerge from this work. First, extending the framework to additional neurological conditions (e.g., epilepsy, dementia) would broaden its clinical utility. Second, incorporating temporal dynamics through recurrent architectures or attention mechanisms could capture time-evolving EEG patterns. Third, investigating transfer learning approaches to address data scarcity in rare neurological conditions represents a valuable research avenue. Finally, prospective clinical validation studies are essential to assess real-world diagnostic performance and clinical impact.

REFERENCES:

[1] Steinmetz JD, Seeher KM, Schiess N, et al. Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet Neurol* 2024; 23: 344–381.

[2] Zhong H, Feng Y, Shen J, et al. Global Burden of Traumatic Brain Injury in 204 Countries and Territories From 1990 to 2021. *Am J Prev Med* 2025; 68: 754–763.

[3] 2018 Guidelines for the Early Management of Patients With Acute Ischemic Stroke: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association | Stroke, https://www.ahajournals.org/doi/10.1161/str.0000000000000158 (accessed 11 September 2025).

[4] Alouani AT, Elfouly T. Traumatic Brain Injury (TBI) Detection: Past, Present, and Future. *Biomedicines* 2022; 10: 2472.

[5] Berrich Y, Guennoun Z. EEG-based epilepsy detection using CNN-SVM and DNN-SVM with feature dimensionality reduction by PCA. *Sci Rep* 2025; 15: 14313.

[6] Caiola M, Babu A, Ye M. EEG classification of traumatic brain injury and stroke from a nonspecific population using neural networks. *PLOS Digit Health* 2023; 2: e0000282.

[7] Modarres MH, Kuzma NN, Kretzmer T, et al. EEG slow waves in traumatic brain injury: Convergent findings in mouse and man. *Neurobiol Sleep Circadian Rhythms* 2017; 2: 59–70.

[8] Sheorajpanday RVA, Nagels G, Weeren AJTM, et al. Quantitative EEG in ischemic stroke: correlation with functional status after 6 months. *Clin Neurophysiol Off J Int Fed Clin Neurophysiol* 2011; 122: 874–883.

[9] van Putten MJAM, Tavy DLJ. Continuous quantitative EEG monitoring in hemispheric stroke patients using the brain symmetry index. *Stroke* 2004; 35: 2489–2492.

[10] Chen W, Liao Y, Dai R, et al. EEG-based emotion recognition using graph convolutional neural network with dual attention mechanism. *Front Comput Neurosci* 2024; 18: 1416494.

[11] Vivaldi N, Caiola M, Solarana K, et al. Evaluating Performance of EEG Data-Driven Machine Learning for Traumatic Brain Injury Classification. *IEEE Trans Biomed Eng* 2021; 68: 3205–3216.

[12] Tong W, Zhang J, Chen F, et al. A novel stroke classification model based on EEG feature fusion. *Sci Rep* 2025; 15: 14287.

[13] Lin P-J, Li W, Zhai X, et al. AM-EEGNet: An advanced multi-input deep learning framework for classifying stroke patient EEG task states. *Neurocomputing* 2024; 585: 127622.

[14] Sawan A, Awad M, Qasrawi R, et al. Hybrid deep learning and metaheuristic model based stroke diagnosis system using electroencephalogram (EEG). *Biomed Signal Process Control* 2024; 87: 105454.

[15] Yuan Y, Xun G, Jia K, et al. A Multi-View Deep Learning Framework for EEG Seizure

Detection. *IEEE J Biomed Health Inform* 2019; 23: 83–94.

[16] Rath A, Mishra D, Panda G, et al. Improved heart disease detection from ECG signal using deep learning based ensemble model. *Sustain Comput Inform Syst* 2022; 35: 100732.

[17] Postepski F, Wojcik GM, Wrobel K, et al. Recurrent and convolutional neural networks in classification of EEG signal for guided imagery and mental workload detection. *Sci Rep* 2025; 15: 10521.

[18] Plewczynski D. Brainstorming: Consensus Learning in Practice. *Front Neuroinformatics*; 3. Epub ahead of print 2009. DOI: 10.3389/conf.neuro.11.2009.08.080.

[19] Thatcher RW. Electroencephalography and Mild Traumatic Brain Injury. In: Slobounov S, Sebastianelli W (eds) *Foundations of Sport-Related Brain Injuries*. Boston, MA: Springer US, pp. 241–265.

[20] Klimesch W. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res Brain Res Rev* 1999; 29: 169–195.

[21] Bazanova OM, Auer T, Sapina EA. On the Efficiency of Individualized Theta/Beta Ratio Neurofeedback Combined with Forehead EMG Training in ADHD Children. *Front Hum Neurosci*; 12. Epub ahead of print 18 January 2018. DOI: 10.3389/fnhum.2018.00003.

[22] Simis M, Imamura M, Pacheco-Barrios K, et al. EEG theta and beta bands as brain oscillations for different knee osteoarthritis phenotypes according to disease severity. *Sci Rep* 2022; 12: 1480.

[23] van Son D, De Blasio FM, Fogarty JS, et al. Frontal EEG theta/beta ratio during mind wandering episodes. *Biol Psychol* 2019; 140: 19–27.

[24] Obeid I, Picone J. The Temple University Hospital EEG Data Corpus. *Front Neurosci*; 10. Epub ahead of print 13 May 2016. DOI: 10.3389/fnins.2016.00196.

[25] Pion-Tonachini L, Kreutz-Delgado K, Makeig S. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage* 2019; 198: 181–197.