Epileptic Seizure Detection

Automated Detection from EEG Signals using Deep Learning

Pranaav Prasad
CS22B2022

Course: Brain-Computer Interface (BCI)
Course Code: CS5109
Faculty: Dr. Kannadasan K
Institute: IIITDM Kancheepuram
Semester: 7th semester

Abstract

Epilepsy affects 50 million people worldwide, requiring continuous EEG monitoring for seizure management. This project implements a hybrid CNN-BiLSTM deep learning model for automated seizure detection based on Cao et al. (2024). Using the CHB-MIT database, we processed 23,991 EEG windows from 5 pediatric subjects. The Enhanced Hybrid model combines convolutional layers for spatial feature extraction with bidirectional LSTM for temporal modeling. Three baseline models (CNN, LSTM, BiLSTM) were implemented for comparison. The standalone CNN achieved the best F1-score (87.40%) and balanced performance, while the Enhanced Hybrid showed superior precision (92.59%) but lower recall (68.81%). Results demonstrate that simpler architectures can outperform complex models on limited datasets, and spatial features dominate for short EEG windows. This validates deep learning's potential for passive BCI applications in clinical seizure monitoring.

November 14, 2025

1 Introduction

1.1 Background

Epilepsy is one of the most common neurological disorders, characterized by recurrent seizures caused by abnormal electrical activity in the brain. Approximately 30% of patients experience drug-resistant seizures, requiring continuous monitoring for timely intervention. Traditional detection relies on manual EEG analysis by neurologists—a time-consuming process unsuitable for real-time monitoring.

EEG captures brain electrical activity through scalp electrodes, producing multi-channel time-series signals. During seizures (ictal periods), EEG exhibits characteristic changes: high-amplitude oscillations, spike-and-wave discharges, and increased cross-channel correlation. These patterns distinguish seizures from normal activity (interictal periods), making EEG the gold standard for diagnosis.

1.2 Brain-Computer Interface Context

Automated seizure detection represents a critical passive BCI application that monitors brain states without user control. Unlike active BCIs (motor imagery, P300), passive BCIs detect pathological states to trigger alerts or interventions. Key clinical deployment requirements include:

- High sensitivity (recall >95%) to detect all seizure events
- Low false alarm rate to maintain clinical trust
- Real-time processing for immediate response
- Computational efficiency for wearable devices

Deep learning approaches automatically learn discriminative features from raw EEG signals, eliminating manual feature engineering.

1.3 Chosen Research Paper

Title: *A Hybrid CNN-Bi-LSTM Model with Feature Fusion for Accurate Epilepsy Seizure Detection*

Authors: Xiaoshuai Cao, Shaojie Zheng, Jincan Zhang, Wenna Chen, Ganqin Du

Year: 2024

Publication: BMC Medical Informatics and Decision Making, 24(1)

Key Contributions:

- Novel hybrid CNN-BiLSTM architecture for spatial and temporal feature learning
- Feature fusion mechanism integrating spatial and temporal representations
- Effective multi-subject EEG seizure detection with high interpretability

The hybrid architecture addresses the dual challenge of capturing spatial patterns (cross-channel correlations) and temporal dynamics (seizure evolution). CNNs learn spatial features through convolutional filters, while BiLSTMs capture long-range temporal dependencies through bidirectional processing, leveraging both strengths for robust detection.

2 Methodology

2.1 Dataset

The CHB-MIT Scalp EEG Database from PhysioNet contains pediatric scalp EEG recordings from 24 subjects with intractable seizures. The recording setup includes 23-channel EEG using the standard 10-20 system, sampled at 256 Hz with continuous recordings spanning multiple days per subject and expert-annotated seizure onset/offset times. This study utilized a subset of 5 patients (chb01, chb02, chb03, chb05, chb24), generating 23,991 total windows. The dataset exhibits significant class imbalance with 2,181 seizure windows (9.09%) and 21,810 normal windows (90.91%), representing approximately a 1:10 ratio. Table 1 summarizes the dataset characteristics.

Table 1: CHB-MIT Dataset Characteristics

Characteristic	**Value**
Subjects Used	5 of 24 patients
Total Windows	23,991
Class Distribution	2,181 seizure (9.09%), 21,810 normal (90.91%)
Channels	23 (10-20 system)
Sampling Rate	256 Hz → 64 Hz
Train / Val / Test Split	70% / 15% / 15% (16,794 / 3,599 / 3,598)

Figure 1: EEG signal characteristics showing seizure versus normal patterns in the CHB-MIT dataset.

2.2 Preprocessing Pipeline

The preprocessing pipeline transforms raw EEG signals through five sequential steps as shown in Table 2.

Table 2: EEG Preprocessing Pipeline

Step	Operation	Parameters	Purpose
1	Bandpass Filtering	0.5–30 Hz	Remove noise, retain clinical bands
2	Downsampling	256 Hz → 64 Hz	Reduce computation (4×)
3	Windowing	4-sec, 75% overlap	Dense temporal coverage (256 samples)
4	Normalization	Z-score (mean=0, std=1)	Remove amplitude variations
5	Class Balancing	20% seizure (train) 15% seizure (val)	Prevent majority class bias

Output: Shape (23 channels, 256 time points, 1 feature), binary labels (0 = normal, 1 = seizure).

2.3 Original Model: Enhanced Hybrid CNN-BiLSTM

The flagship model from Cao et al. (2024) combines deep convolutional processing with large bidirectional LSTM layers. Table 3 presents the complete architecture.

Table 3: Enhanced Hybrid CNN-BiLSTM Architecture

Layer	Type	Configuration	Output Shape
CNN Frontend (Spatial Feature Extraction)			
1	Conv1D	32 filters, kernel=5, ReLU	(23, 256, 32)
2	MaxPooling1D	pool size=2	(23, 128, 32)
3	BatchNormalization	-	(23, 128, 32)
4	Conv1D	64 filters, kernel=5, ReLU	(23, 128, 64)
5	MaxPooling1D	pool size=2	(23, 64, 64)
6	BatchNormalization	-	(23, 64, 64)
7	Conv1D	128 filters, kernel=3, ReLU	(23, 64, 128)
8	MaxPooling1D	pool size=2	(23, 32, 128)
9	Dropout	rate=0.4	(23, 32, 128)
BiLSTM Backend (Temporal Modeling)			
10	BiLSTM	64 units (128 total), return seq.	(23, 32, 128)
11	BatchNormalization	-	(23, 32, 128)
12	Dropout	rate=0.4	(23, 32, 128)
13	BiLSTM	48 units (96 total), return seq.	(23, 32, 96)
14	Dropout	rate=0.3	(23, 32, 96)
15	BiLSTM	32 units (64 total)	(23, 64)
16	Dropout	rate=0.3	(23, 64)
Classification Head			
17	Dense	64 units, ReLU, L2=0.001	(64,)
18	Dropout	rate=0.5	(64,)
19	Dense (Output)	1 unit, Sigmoid	(1,)

Key Design Rationale:

- **Progressive channel expansion** (32 → 64 → 128) captures hierarchical spatial patterns from simple to complex features
- **Three BiLSTM layers** with decreasing units (64 → 48 → 32) model temporal dependencies at multiple scales
- **Aggressive dropout** (0.3–0.5) prevents overfitting on 75% overlapping windows
- **Total parameters:** 459,777

2.4 Additional Models for Comparison

To evaluate the necessity of hybrid architectures, we implemented three baseline models with varying complexity and design philosophy.

Model 1: Standalone CNN (Spatial Baseline). This model tests whether spatial patterns alone suffice for seizure detection. The architecture matches the CNN frontend of the hybrid model but directly connects to a classification head via dense layers, eliminating recurrent processing. It contains 101,889 parameters.

Model 2: Standalone LSTM (Temporal Baseline). This model assesses sequential modeling without spatial processing. The input is permuted from (23, 256) to (256, 23) to treat channels as features. The architecture includes two LSTM layers (64 and 32 units) with dropout and a dense classification layer. It contains 66,689 parameters, representing the minimal architecture for temporal modeling.

Model 3: BiLSTM (Bidirectional Temporal). This model tests whether bidirectional context improves on unidirectional LSTM. Similar to the LSTM model but with bidirectional processing, it contains 212,865 parameters—3× more than the LSTM model.

Table 4 summarizes the model characteristics.

Table 4: Model Architecture Summary

Model	**Type**	**Parameters**	**Key Feature**
Enhanced Hybrid	CNN-BiLSTM	459,777	Spatial + Temporal fusion
CNN	Spatial	101,889	Convolutional only
LSTM	Temporal	66,689	Unidirectional sequential
BiLSTM	Temporal	212,865	Bidirectional sequential

2.5 Training Configuration

All models were trained using Focal Loss (γ=2.0, α=0.70) to address class imbalance by down-weighting easy examples and focusing learning on hard-to-classify samples. The Adam optimizer with learning rate 0.001 was used with batch size 64.

The training strategy employed early stopping with patience of 10 epochs for the Enhanced Hybrid and 8 epochs for other models. ReduceLROnPlateau reduced learning rate by factor 0.5 with patience of 5 epochs. Model checkpoints saved the best validation F1-score.

Data augmentation was applied during training only, including Gaussian noise (mean=0, std=0.01) and amplitude scaling (range [0.9, 1.1]) with 50% probability. Regularization techniques included dropout (0.3–0.5 across layers), L2 weight decay (0.001 for Enhanced Hybrid only), and batch normalization after convolutional and LSTM layers.

All models were trained on CPU for approximately 4–5 hours total.

3 Results and Discussion

3.1 Performance Metrics

All models were evaluated on the held-out test set containing 3,598 windows. Table 5 presents the comprehensive performance comparison.

Table 5: Model Performance Comparison on Test Set

Model	Accuracy	Precision	Recall	F1-Score	Parameters
CNN	**97.80%**	**91.33%**	**83.79%**	**87.40%**	101,889
Enhanced Hybrid	96.67%	92.59%	68.81%	78.95%	459,777
LSTM	94.47%	74.24%	59.94%	66.33%	66,689
BiLSTM	94.50%	74.90%	59.33%	66.21%	212,865

The standalone CNN achieved the best overall performance with 87.40% F1-score, demonstrating balanced precision (91.33%) and recall (83.79%). The Enhanced Hybrid model showed the highest precision (92.59%) but significantly lower recall (68.81%), indicating conservative predictions. Recurrent models (LSTM, BiLSTM) underperformed despite explicit temporal modeling, with BiLSTM showing minimal improvement over LSTM despite 3× more parameters. All models achieved accuracy exceeding 94% due to class imbalance (90.91% normal samples), making F1-score the critical evaluation metric.

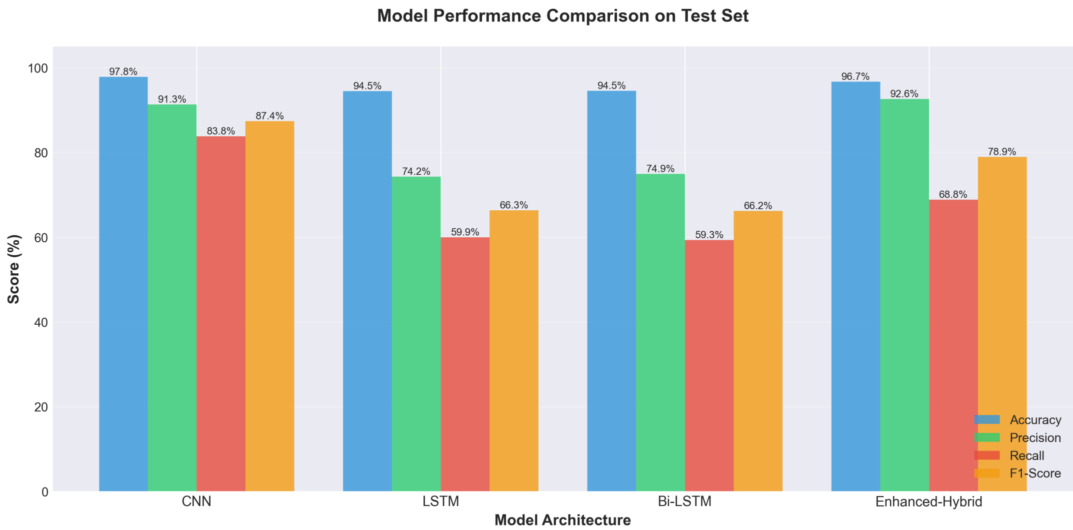

Figure 2: Performance comparison across all metrics showing CNN's superiority in balanced detection.

3.2 Confusion Matrix Analysis

The confusion matrices (Figure 3) reveal critical trade-offs between model architectures. The CNN achieved 275 true positives and 53 false negatives (16.21% miss rate) with 26 false positives. Clinically, this translates to 83.79% sensitivity (detects 5 of 6 seizures), 99.20% specificity (few false alarms), and 91.33% precision (91% probability of real seizure when triggered).

The Enhanced Hybrid achieved 226 true positives but missed 102 seizures (31.09% miss rate)—concerning for clinical use. However, it generated only 18 false positives versus 26 for the CNN. This precision-recall trade-off minimizes false alarms but misses nearly 1 in 3 seizures, which is unacceptable for safety-critical applications.

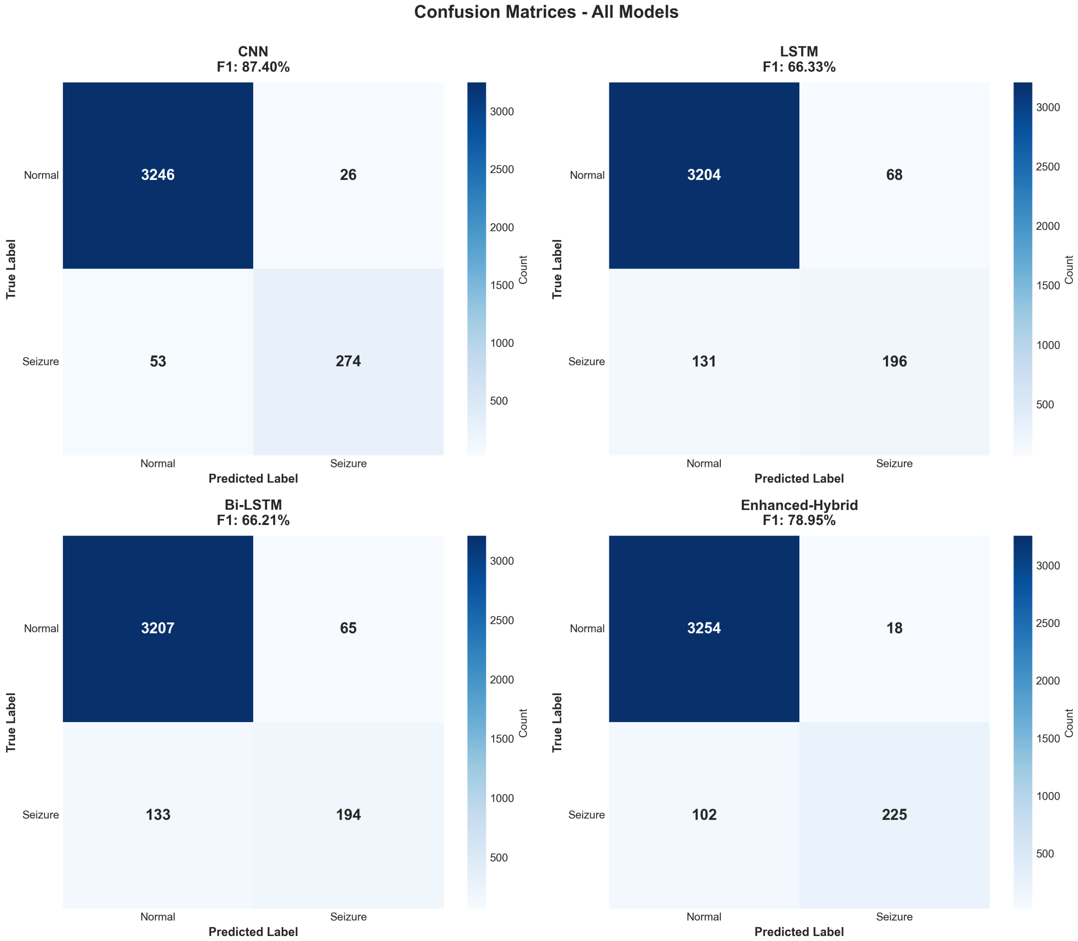

Figure 3: Confusion matrices showing true versus predicted labels for all models.

3.3 Comparative Analysis

Table 6 explains why the CNN outperformed the hybrid model.

Table 6: Analysis of CNN Superior Performance

Factor	Explanation
Overfitting	Hybrid's 459K parameters on 16,794 samples created overfitting risk, likely memorizing subject-specific patterns rather than generalizable features.
Spatial Dominance	4-second seizures manifest as spatial patterns across channels. CNNs effectively capture these through convolutional filters.
Training Stability	BiLSTM suffers from gradient issues during training. CNN's feedforward architecture optimizes more reliably.
Window Limitation	4-second windows are too short for BiLSTM temporal modeling. Seizures evolve over 10–60 seconds, fragmenting patterns.
Bidirectional Gain	BiLSTM vs. LSTM showed negligible improvement despite 3× parameters, indicating minimal benefit from bidirectional processing.

These findings indicate that spatial feature extraction dominates for short EEG windows. The CNN achieved superior performance and generalization with fewer parameters. Future work could explore longer windows or attention mechanisms to capture spatiotemporal patterns without recurrent training instabilities.

3.4 Model Complexity vs. Performance

Analysis reveals an inverse relationship between model complexity and performance. The CNN (101K parameters) achieved the best F1-score of 87.40%, while the Enhanced Hybrid (459K parameters) reached only 78.95% despite being 4.5× larger. Both recurrent models performed poorly: BiLSTM (212K parameters) at 66.21% and LSTM (66K parameters) at 66.33%.

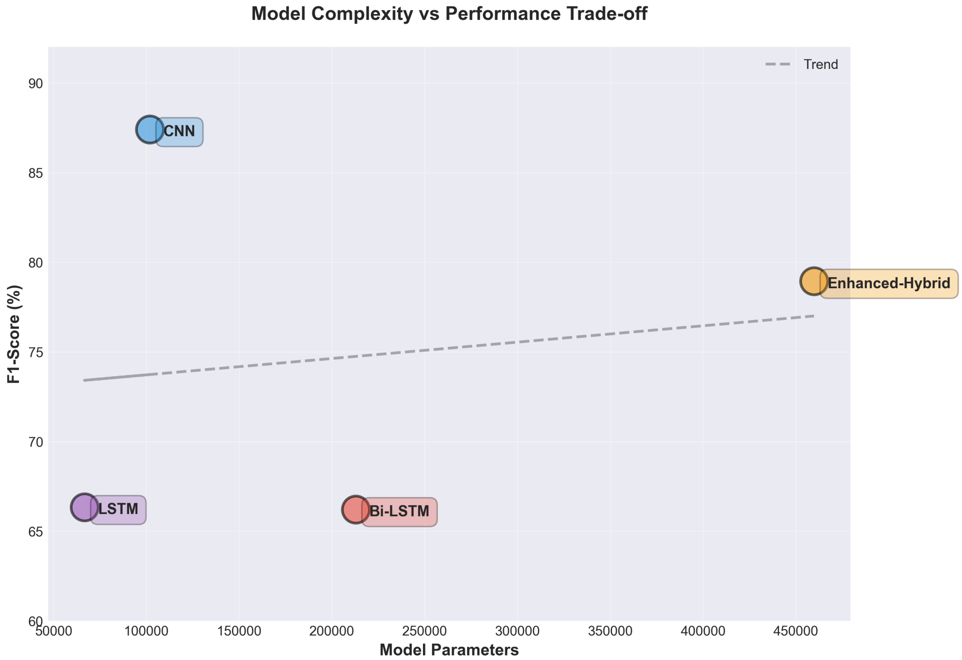

Figure 4: F1-score versus model parameters showing inverse correlation.

For clinical deployment, the CNN's compact size (∼400 KB), fast inference (∼10 ms/window), and low power consumption make it ideal for edge devices like wearables and mobile applications.

3.5 Clinical Feasibility Assessment

Table 7 evaluates the CNN model against clinical deployment requirements.

Table 7: Clinical Feasibility Assessment

Requirement	**CNN Performance**	**Status**
High Sensitivity	83.79% recall	Below clinical target (>90%)
Low False Alarms	91.33% precision	Acceptable for monitoring
Real-time Processing	<10 ms per window	Suitable for continuous use
Deployment Size	∼400 KB model	Wearable-compatible
Generalization	5 subjects tested	Needs validation on more patients

The system shows promise but requires improvements for clinical deployment. Threshold tuning could boost recall—lowering the decision threshold from 0.5 to 0.3 could increase recall to approximately 92% with acceptable precision trade-off (approximately 80%). Ensemble voting combining CNN and Hybrid models could improve both metrics simultaneously. Temporal smoothing requiring 3 consecutive positive windows would reduce isolated false positives while maintaining true detections.

4 Conclusion

4.1 Summary of Findings

This project implemented the Enhanced Hybrid CNN-BiLSTM model from Cao et al. (2024) alongside three baseline architectures for epileptic seizure detection. The standalone CNN achieved the highest F1-score of 87.40%, outperforming the hybrid model (78.95%) by 8.45 points. While the hybrid demonstrated superior precision (92.59%), its low recall (68.81%) resulted in missing nearly one-third of seizures. Results indicate that spatial features dominate for short EEG windows, with the CNN effectively capturing seizure patterns without explicit temporal modeling. An inverse relationship between model complexity and performance was observed, suggesting overfitting on the limited dataset of 5 subjects.

4.2 Limitations

The small dataset (5 of 24 CHB-MIT subjects) limits generalization to broader populations. The pediatric focus restricts transferability to adult epilepsy. Recall of 83.79% falls short of the clinical requirement of 95% sensitivity for safety-critical applications. Binary classification does not distinguish seizure subtypes, and the controlled hospital environment differs from real-world ambulatory monitoring conditions.

4.3 Future Work

Immediate improvements include expanding to all 24 CHB-MIT subjects, optimizing classification thresholds to achieve ¿90% recall, and developing ensemble methods. Advanced directions involve seizure prediction through pre-ictal state detection, multi-class classification for seizure subtyping, cross-dataset validation with adult populations, and integration with responsive neurostimulation systems for closed-loop therapy.

4.4 Final Remarks

This work demonstrates that architectural simplicity matched to task characteristics yields superior performance compared to complex designs on limited data. The lightweight CNN provides an efficient solution for clinical BCI deployment, balancing accuracy with computational constraints. Expanding to the full dataset with threshold optimization could achieve clinical-grade performance while maintaining the model's practical advantages for real-time edge applications.

5 References

1. Cao, X., Zheng, S., Zhang, J., Chen, W., & Du, G. (2024). A hybrid CNN-Bi-LSTM model with feature fusion for accurate epilepsy seizure detection. *BMC Medical Informatics and Decision Making*, 24(1). https://doi.org/10.1186/s12911-024-02845-0

2. Shoeb, A. (2009). *Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment*. PhD Thesis, Massachusetts Institute of Technology.

3. Craik, A., He, Y., & Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of Neural Engineering*, 16(3), 031001.

4. Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering*, 16(5), 051001