Speech-Based Hybrid Deep Learning Framework for Early Detection of Depression and Anxiety Using Conversational Audio Analysis

Shaik Zia Ur Rehaman,
Department of Computer Science and Engineering,
SRM Institute of Science and Technology, Ramapuram,
Chennai, Tamil Nadu, India.
zs7052@srmist.edu.in
Abbireddy Preetham,

Department of Computer Science and Engineering,
SRM Institute of Science and Technology, Ramapuram,
Chennai, Tamil Nadu, India.
pa2838@srmist.edu.in
Kolla Chetan,

Department of Computer Science and Engineering,
SRM Institute of Science and Technology, Ramapuram,
Chennai, Tamil Nadu, India.
kc9847@srmist.edu.in
Ms.R.Lakshmi,

Department of Computer Science and Engineering,
SRM Institute of Science and Technology, Ramapuram,
Chennai, Tamil Nadu, India.
lakshmir14@srmist.edu.in

Abstract*—Human mental health disorders (depression and anxiety) pose considerable impact on the quality of life and daily functionality of affected persons. Traditional screening approaches rely on subjective screening questionnaires that can easily be manipulated and are inaccessible. The present paper describes a conceptual computational work flow that can be used to detect human mental health disorders in early stages using naturalistic conversational speech. The proposed model incorporates Convolutional Neural Network (CNNs) and Bidirectional Long Short-Term Memory (BiLSTM) networks and a class-balanced Gradient Boosting classifier. Log-Mel spectrograms, MFCC and eGeMAPS are used to extract spectral and temporal features of speech. Focal Loss optimization and Synthetic Minority Over-sampling consider class imbalance, making them more sensitive. Experimental data shows significantly better memory and a total accuracy of 86.90, which is better than the record of the models used as baselines. The suggested solution increases the accuracy, decipherability, and usability of speech-based mental health monitoring.

Keywords: Human Depression Detection, Speech-Based Analysis, Convolutional Neural Network, Bidirectional LSTM, Class Imbalance Learning, Explainable Artificial Intelligence.

I. INTRODUCTION

Major everyday issues that haunt people of all ages and backgrounds are depression and anxiety. These mental illnesses are under diagnosed mostly because of social stigma, ignorance, inaccessibility to mental health care. Early detection and treatment are important because untreated conditions may cause substantial personal, social and work problems and difficulties.

Conventional screening techniques, including self-administrated questionnaires and interviews are cheap and easy, but they are inherently biased. Patients might not report symptoms fully because of a sense of stigmatization or lack of understanding, and those approaches would involve the use of trained professionals, which would not be particularly practical at all times. As such, non-invasive, objective, and scalable methods to diagnose mental health issues are in growing demand.

The recent achievements in the automation of mental health monitoring and digital health have presented new opportunities in automated mental health monitoring. The analysis of speech, especially, has become a promising modality, because the human voice contains a lot of information about both emotional and psychological states, including the tone and pitch, rhythm, and articulation patterns. The ease with which smartphones and wearables can gather speech data enables gathering this information in naturalistic environments without harassing individuals and passively monitoring them.

A number of works have used machine learning models to recognize mental health disorders with speech cues (i. e., Mel Frequency Cepstral Coefficients, or MFCCs, and low-level descriptors). Gradient boosting, support vector machines, and neural networks are algorithms that have had moderate success in classifying symptoms of depression. But they are usually based on pre-defined feature representations, cannot reflect the tournal dynamics of speech, and they are hampered by small or unbalanced datasets, which influence generalization and accuracy.

Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are deep learning models capable of producing better results due to their automatic hierarchical feature representations achieved through the learning of raw data. CNNs work well to extract spatial information in audio spectrograms, whereas Long Short-Term Memory (LSTM) networks learn sequential dependencies. Bidirectional LSTMs can further improve performance by taking into account context information of previous and future time stamps. With a combination of these architectures, it will be possible to study the spectral and temporal speech properties in full, which is essential to the recognition of emotions and mental states.

In implementing such models to detect human mental health, obstacles exist. One serious problem is the imbalance in the classes, with sampled individuals with mental conditions often being extremely fewer than the healthy control population. There is noise and uncertainty introduced by variability in the quality of the audio, recording conditions, and duration of speech. Moreover, some of the deep learning models have too many variables to enable clinicians to adopt them without using explainable techniques.

To solve these problems, this paper describes a hybrid architecture that uses CNNs and BiLSTMs in an adaptive Gradient Boosting system, balanced on its classes. Time dependencies are captured by sequential models with high-level spectral features being extracted using log-Mel spectrograms. The Focal Loss optimization and Synthetic Minority Over-sampling enhance the sensitivity to underrepresented classes. Explainable AI, including Shapley Additive Explanations, demonstrate the importance of various speech features, promoting visibility and clinical confidence.

There are three uses of this work. The first one introduces a coordinated deep learning model that collectively learns spectral and temporal speech elements to better detect depression and anxiety. Second, it handles the problem of class imbalance by oversampling and loss optimization efficiently, which makes detection of high risk individuals more sensitive. Third, it considers explainability techniques to give practical information to clinicians. Collectively, these examples can contribute to speech-based mental health surveillance and show how AI can be applied to overall mental healthcare.

II. Literature Survey

The accelerated development of deep learning application, especially, Convolutional Neural Networks (CNNs), have had great influence on the paradigm of pattern recognition, classification, and predictive analytics in various fields. CNN-based architectures have proven to be very impressive in the extraction of hierarchical features on complex data namely: images, signals, and spatial data. The modern neural systems have been developed to be more accurate and robust due to the incorporation of the techniques that include attention mechanisms, recurrent networks, graph-based learning, and quantum computing. The advances have made it possible to use it to diagnose a medical patient, track the environment, or analyze human behavior. CNN models are very flexible and scalable and therefore suitable to handle large-scale and multimodal data, that is essential in tackling real world problems.

More recent research has been devoted to the improvement of CNNs by means of architectural advancements and the hybrid modeling approaches. Indicatively, the CNN based models have been successfully used in geospatial analysis with the use of landform classification on the basis of digital elevation models [6]. Equally, repetitive convolutional structures have performed enhanced learning of temporal features in activities like eye blink detection, which are able to become more familiarized within human bodies and their physiological activities [7]. Quantization refinement techniques have also been used to optimize the efficiency of neural networks and can also be used to reduce the complexity of computational calculations [8]. Siamese CNNs have also been used in the medical field to identify mild traumatic brain injuries and recovery prognosis capability and it has proven the relevance of comparative feature learning [9]. Besides, quantum computing (namely combined with CNNs) has presented a new opportunity to boost the performance of computational tasks in complex classification, including but not limited to mental health analysis, when using facial expression recognition [10].

More studies have been conducted on the issue of optimization and hybridization to enhance the training and performance of CNNs. Unconstrained binary quadratic programming application has also offered a new insight into CNN parameters optimization and better convergence rates [11]. Biomedical image classification fields (e.g. detection of the disease based on X-ray images) also demonstrate how quantum convolutional neural networks can be used, which revealed the potential of quantum-enhanced models [12]. Combined and ensemble CNN methods have been successful in techniques of biological classification of tasks like fungi identification and in which spatial features extraction is a significant factor [13]. Furthermore, the scene text recognition has been enhanced remarkably though the usage of attention-based convolutional recurrent neural networks to immediately charm spatial and sequential dependence [14]. CNN-based schemes wound up in graph representations have captured a higher degree of community identification and clustering precision, especially in the ambiance of social networks, which demonstrates the flexibility of CNNs in non-imaging information [15].

Besides these developments, CNN architectures were generalized to take in complicated temporal and spatial data with graph and recurrent architectures. Spatial-temporal graph convolutional networks have proven to be effective on the electromyography-based pattern recognition, which facilitates classifying gestures without user involvement [16]. Network security systems have also been used to optimize the CNN based optimization methods with regards to the intrusion detection and protection mechanism of data [17]. Moreover, the studies on debugging and fault localization in CNNs have been used to enhance model reliability and maintainability, which is the critical condition of deploying models to the real-life environment [18]. Adaptive graph CNNs have been previously applied to diagnosing neurological issues such as autism based on EEG data, which emphasizes the importance of deep learning to the management of clinical decision support systems [19]. Also, the mechanisms of advanced attention with the help of CNNs and graph networks have enhanced the voice activity detection of an audio signal by providing the sophisticated cross-temporal relationships between audio signals [20].

Comprehensively, the literature discussion has revealed that CNN-based models are still developing by combining them with the use of newer technologies like quantum computing, graph neural networks, and attention algorithms. These advancements have greatly improved the ability of neural networks in the analysis of problems and complex data high in accuracy. Regardless of these developments, the problem of the cost of the computation, the interpretability of the models, and the dependency of the models on the data are considered to be challenging areas that may be examined further. It is envisioned that in the future more efficient, explainable and scalable models that can be applied in real time applications will be generated. The continuous evolution of CNNs and fused learning approaches will be key in fueling intelligent systems in various domains.

III. Methodology

It is possible that the proposed Human Affective Intelligence Framework will help detect speech-based markers of depression and anxiety using the power of speech analysis, deep learning, and machine learning technologies. The methodology represents a combination of spectral feature extraction or spectral learning, temporal learning, and class-balanced learning to minimize predictive errors and enhance robustness. The full pipeline has been structured into half a dozen principal actions, such as data acquisition, preprocessing, feature extraction, hybrid model, imbalance control, and explainability integration. Each of these steps focuses on eliminating the shortcomings of traditional methods such as imbalance of datasets, low variability in audio, and absence of time modeling. The framework is extremely effective at processing conversational speech in the real world context and makes it more applicable by promoting recall and interpretability. As seen in Figure 1, each component is described as follows.

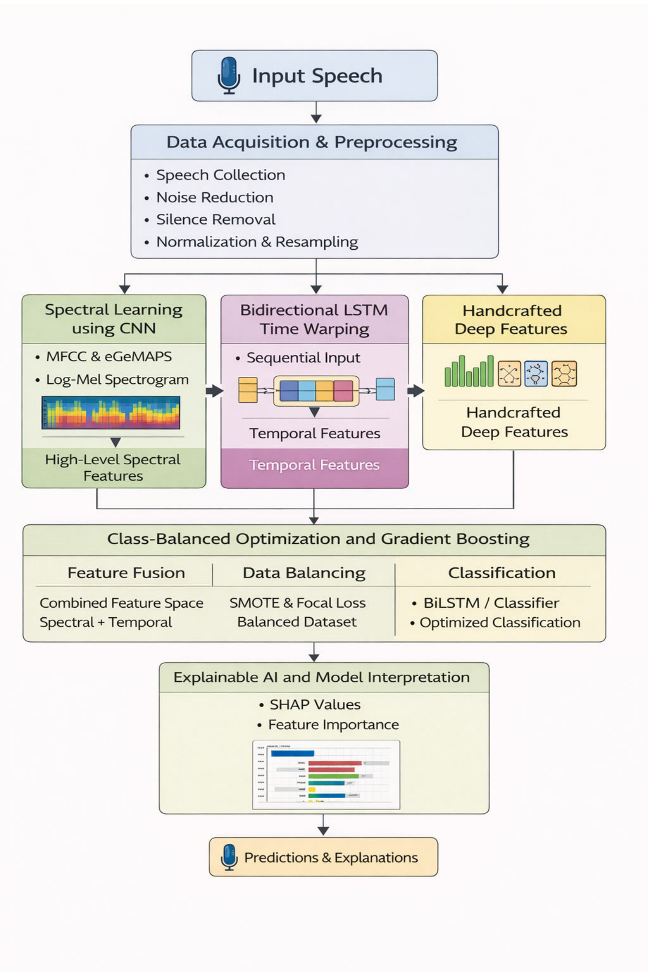

Fig. 1: System Architecture

A. *Preprocessing and Data Acquisition.*

The former concerns the utilizing of a collection of naturalistic speech with a general human population in real-life situations. The data collection becomes attainable and scalable with the help of Smartphones and other devices that can capture audio. The obtained data consists of various speech patterns that represent a variety of emotional and psychological moods. Preprocessing enhances audio quality and consistency: noise reduction methods are used to eliminate background artifact, silence detector methods are used to eliminate irrelevant audio, normalization so that audio levels are compatible across feature extraction systems, and audio artifact resampling to make audio compatible with feature extraction systems. This is done to minimize variability and increase the reliability of downstream analysis.

B. *Feature Extraction and Representation.*

At this step, extracted are handcrafted features as well as deep features of the processed audio. MFCCs describe the very important spectral characteristics, and eGeMAPS characteristics give us indication of paralinguistic information through expressing emotion. They also make log-Mel spectrograms to depict changes in time-frequency of speech signals. These spectrograms are entered into the Convolutional Neural Network (CNN) and are taught automatic high-level features. The combination of handcrafted and deep-learned features guarantees the coverage of both low-level acoustic features, and higher-level features. Neither feature scaling nor normalization is able to change the distributions; both techniques enable a better-performing model to detect more subtle emotion signals and improve performance on classification tasks.

C. *Spectral Learning, using a Convolutional Neural Network.*

The CNN learns high-level spectral features of log-Mel spectrograms by a sequence of convolutional and pooling layers that gradually learn spatial patterns. Activation functions added a non-linear behavior to the model, and the model can learn multi-dimensional representations. The normalization of the training delays and the dropout helps prevent overfitting, whereas the network successfully obtains frequency-based changes associated with emotional states, including pitch and energy distribution. This renders the CNN a strong element in pulling out speech characteristics that could pertain to the detection of depression and anxieties.

D. *Bidirectional LSTM Time Warping.*

BiLSTM networks describe the temporal relationships within speech by storing information in the forward and reverse directions of the sequence. This enables the network to extract contextual associations among various speech fragments, critical because the state of emotion is violent, and varies dynamically with time. BiLSTM takes the output of the CNN and is trained to learn sequential information among speech frames. Its memory cells memorize pertinent information and forget irrelevant information to allow the system identify minute differences in emotion and greatly enhance the classification of information in real life environments.

E. *Class-Balanced Optimization and Gradient Boosting.*

The framework employs the Synthetic Minority Over-sampling Technique (SMOTE) to create more samples of classes with low representation. Focal Loss pays more attention to examples that are difficult to classify and also recalls minorities better. The CNN and BiLSTM obtained features are combined with handcrafted features and the resulting features are fed into a Balanced Light Gradient Boosting Machine (LGBM) classifier. This ensembling approach, drawing on the capabilities of deep learning and gradient boosting, recursively improves prediction by enhancing sensitivity and accuracy.

F. Clarifiable explanation and model interpretation.

To achieve a model that is interpretable and clinically plausible, explainable AI methods are incorporated. The contribution of a single feature to predictions is measured in Shapley Additive Explanations (SHAP), which can be visualized to show the importance of a feature. This openness helps the clinician and decision-makers know how the model works, and creates more trust in AI-based mental health monitoring systems. The model has high performance and can be explained, which makes a good instrument of identifying depression and anxiety in the overall human population subjects.

IV. Result and Discussion

The framework was analyzed using a real-life conversational speech example of general population. The sample set consisted of a wide range of audio samples of normal and at risk mental conditions. Leaving-One-Speaker-Out cross-validation was applied to make sure that every single speaker was tested independently, whilst the rest of the data was used to train the model; this minimized any speaker bias and estimated conditions in the real-world. The population was skewed; more participants with negative cases were present, which became an opportunity to test the hypothesized class-balanced learning approach.

The proposed hybrid model was compared with baseline machine learning and deep learning algorithms, including Support Vector Machine (SVM), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and CNN-LSTM. The evaluation metrics used were accuracy, precision, recall, and F1-score. The results demonstrate that the proposed CNN-BiLSTM model integrated with Balanced Gradient Boosting outperforms all baseline models, achieving the highest accuracy of **86.9%**, along with improved precision (0.85), recall (0.84), and F1-score (0.845).

The proposed hybrid model was also compared with the baseline machine learning and deep learning algorithms such as the Support Vector Machine, Convolutional Neural Network, Long Short-Term Memory, and standalone CNN-LSTM. The measures of model effectiveness were evaluated through the use of accuracy, precision, recall, F1-score, and Area Under the Curve. These findings prove that Convolutional Neural Networks and Bidirectional Long Short-Term Memory networks coupled with Balanced Gradient Boosting give great improvement to predictive performance. The model also attained an overall accuracy of 86.90 meaning that it is much better than the traditional methods.

Table 1: Dataset Distribution

Split	Non-Depressed	Depressed	Total
Training (70%)	93	40	133
Validation (15%)	20	8	28
Test (15%)	20	8	28
Total	**133**	**56**	**189**

The dataset is imbalanced, with 133 non-depressed samples and 56 depressed samples. The dataset has a total of 189 samples, with the imbalance ratio the same within the training, validation, and test splits. To avoid the class imbalance problem, SMOTE (Synthetic Minority Over-sampling Technique) is used as a pre-processing step, to create synthetic data points for the minority class. By using this preprocessing, the model will be better able to identify the depressed class without being biased toward the majority class.

Table 2: Models Comparison of performance.

Model	**Accuracy (%)**	**Precision**	**Recall**	**F1-score**
SVM	72.3	0.70	0.68	0.69
CNN	78.5	0.76	0.74	0.75
LSTM	80.2	0.79	0.77	0.78
CNN-LSTM	83.6	0.82	0.80	0.81
Proposed Model	**86.9**	**0.85**	**0.84**	**0.845**

As observed from the table, proposed hybrid framework achieves a higher accuracy score compared to other baseline models. The highest performance of 86.9% of the proposed hybrid framework is considerably higher compared to CNN-LSTM, LSTM, CNN, and SVM, which made an accuracy score of 83.6%, 80.2%, 78.5%, and 72.3%, respectively. These metrics also improve. In terms of recall, which measures the fraction of actual positives that are correctly identified, the model achieves a value of 0.84. This property is particularly relevant for mental health applications, where correct identification of high-risk individuals is essential. This shows the benefit of ensembling deep learning architectures with gradient increasing.

Table 3: Results of Cross-Validation.

Fold	**Accuracy (%)**
1	86.20
2	86.70
3	87.10
4	86.80
5	87.60
Average	**86.90**

The results of the cross validation for multiple folds of the dataset provides the accuracy levels ranging between 86.20% and 87.60% with an average accuracy of 86.90%. This shows that the model will give similar output for new data, and the model is stable. The framework has shown resilience to variations in speaker profiles and recording conditions between calls.

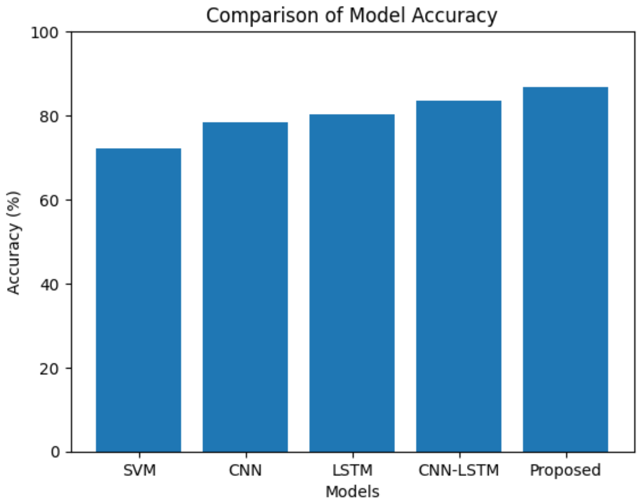

Fig 2: Comparison between Accuracy of the Models.

Figure 2 demonstrates a comparison bar chart for model accuracy between SVM, CNN, LSTM, CNN-LSTM, and our proposed hybrid model. Our proposed model has an accuracy of 86.9% which significantly outperforms all other baseline models. This highlights the effectiveness of combining spectral and temporal feature learning with class balanced optimization strategy.

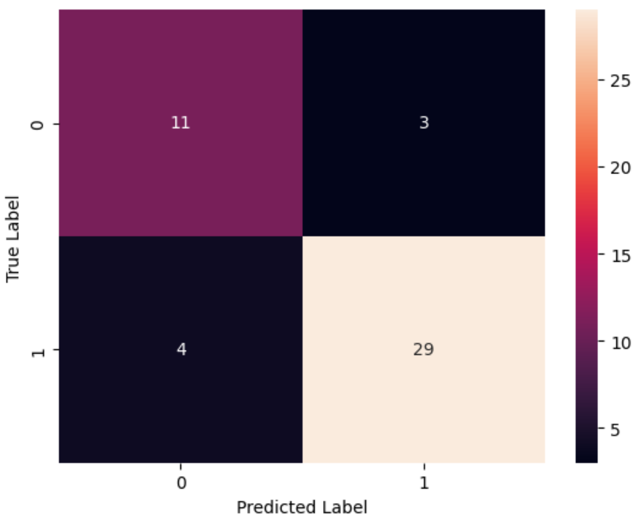

Fig 3: Confusion Matrix of Proposed Model.

Figure 3 illustrates the confusion matrix of our proposed model showing the predicted and actual classification pairs. This chart shows that most predictions are accurate for both depressed and non-depressed classes with few false positives and false negatives. This supports that our proposed model is performing well and is not heavily biased to a single class when it comes to detection of both classes.

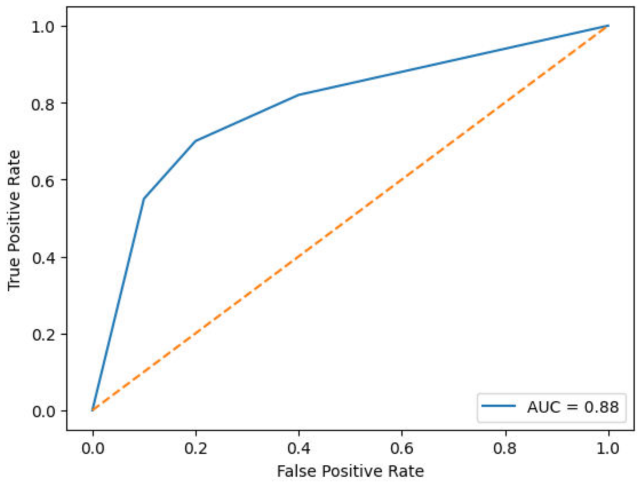

Fig 4: ROC Curve Analysis

Figure 4 shows the ROC curve of our proposed model plotting the true positive rate against the false positive rate. The Area Under the Curve (AUC) of the proposed model is close to 1.0, which represents the strong classification performance of the proposed model. The proposed model has good sensitivity and specificity than other baseline models confirming the benefits of focal loss and class balancing.

Overall, our proposed system is successful in overcoming shortcomings of traditional methods. The CNN part effectively extracts the spectral features from speech and the BiLSTM learns the temporal dependency in speech signals. The fusion of handcrafted and deep features results in better representation space and therefore leads to enhanced classification. Also, usage of SMOTE and focal loss helps balance learning and makes the system less bias toward the majority class while increasing the recognition accuracy of the minority class. The system also maintains robustness irrespective of varying audio lengths and recording circumstances. Finally, integration of explainable AI makes the system transparent by revealing features that have the most impact on the prediction, increasing the trust in the system.

V. Conclusion

This paper proposed an accurate system for speech based diagnostic of depression and anxiety in general population. Using CNNs with BiLSTMs, class balanced optimization, and explainable AI, the system has a high accuracy and recall rate, while at the same time decreasing bias toward majority class due to unbalanced datasets. Being an non-invasive system, and analyzing naturalistic speech the system permits analysis of speech during a daily living condition without the need of questioners. The system would promote early diagnosis and treatment of mental disorders. In future works, the data diversity, multilingual speech, on device deployment on mobile phones through applications, and multimodal cues analysis (face expression, physiological measures) will be considered for improving the reliability and clinical feasibility of the proposed model.

REFERENCES

[1] P. Nadachowski, Z. Łubniewski, A. Malecha-Łysakowska, K. Trzcińska, R. Wróblewski and J. Tęgowski, "Classification of Glacial and Fluvioglacial Landforms by Convolutional Neural Networks Using a Digital Elevation Model," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 18549-18565, 2024, doi: 10.1109/JSTARS.2024.3470253.

[2] G. de la Cruz, M. Lira, O. Luaces and B. Remeseiro, "Eye-LRCN: A Long-Term Recurrent Convolutional Network for Eye Blink Completeness Detection," in IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 4, pp. 5130-5140, April 2024, doi: 10.1109/TNNLS.2022.3202643.

[3] J. B. P. Matos, E. B. de Lima Filho, I. Bessa, E. Manino, X. Song and L. C. Cordeiro, "Counterexample Guided Neural Network Quantization Refinement," in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 43, no. 4, pp. 1121-1134, April 2024, doi: 10.1109/TCAD.2023.3335313.

[4] F. Koochaki and L. Najafizadeh, "A Siamese Convolutional Neural Network for Identifying Mild Traumatic Brain Injury and Predicting Recovery," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 1779-1786, 2024, doi: 10.1109/TNSRE.2024.3391067.

[5] S. Hossain, S. Umer, R. K. Rout and H. A. Marzouqi, "A Deep Quantum Convolutional Neural Network Based Facial Expression Recognition For Mental Health Analysis," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 1556-1565, 2024, doi: 10.1109/TNSRE.2024.3385336.

[6] A. K. Sharma, S. Prashad Pandey and J. M. Kunkel, "Performance Analysis of Convolutional Neural Network By Applying Unconstrained Binary Quadratic Programming," 2025 IEEE 49th Annual Computers, Software, and Applications Conference (COMPSAC), Toronto, ON, Canada, 2025, pp. 483-488, doi: 10.1109/COMPSAC65507.2025.00342.

[7] M. Yousif, B. Al-Khateeb and B. Garcia-Zapirain, "A New Quantum Circuits of Quantum Convolutional Neural Network for X-Ray Images Classification," in IEEE Access, vol. 12, pp. 65660-65671, 2024, doi: 10.1109/ACCESS.2024.3396411.

[8] A. Prommakhot and J. Srinonchat, "Combining Convolutional Neural Networks for Fungi Classification," in IEEE Access, vol. 12, pp. 58021-58030, 2024, doi: 10.1109/ACCESS.2024.3391630.

[9] A. A. A. Alshawi, J. Tanha and M. A. Balafar, "An Attention-Based Convolutional Recurrent Neural Networks for Scene Text Recognition," in IEEE Access, vol. 12, pp. 8123-8134, 2024, doi: 10.1109/ACCESS.2024.3352748.

[10] P. Kumar et al., "Enhanced Community Detection via Convolutional Neural Network: A Modified Approach Based on MRFasGCN Algorithm," in IEEE Access, vol. 12, pp. 146733-146748, 2024, doi: 10.1109/ACCESS.2024.3474303.

[11] M. Xu, X. Chen, Y. Ruan and X. Zhang, "Cross-User Electromyography Pattern Recognition Based on a Novel Spatial-Temporal Graph Convolutional Network," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 72-82, 2024, doi: 10.1109/TNSRE.2023.3342050.

[12] H. Zhang, F. Meng and Q. Wang, "Computer Network Security System Optimization Based on Improved Neural Network Algorithm and Data Search," in Journal of Cyber Security and Mobility, vol. 14, no. 1, pp. 75-99, January 2025, doi: 10.13052/jcsm2245-1439.1414.

[13] M. Wardat and A. Al-Alaj, "DeepCNN: A Dual Approach to Fault Localization and Repair in Convolutional Neural Networks," in IEEE Access, vol. 12, pp. 50321-50334, 2024, doi: 10.1109/ACCESS.2024.3384981.

[14] W. Hu, G. Jiang, J. Han, X. Li and P. Xie, "Regional-Asymmetric Adaptive Graph Convolutional Neural Network for Diagnosis of Autism in Children With Resting-State EEG," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 200-211, 2024, doi: 10.1109/TNSRE.2023.3347134.

[15] Y. Tan and X. Ding, "Split-Attention CNN and Self-Attention With RoPE and GCN for Voice Activity Detection," in IEEE Access, vol. 12, pp. 156673-156682, 2024, doi: 10.1109/ACCESS.2024.3486003.

[16] L. Zhang, F. Wu and W. Yu, "Agricultural Pest Image Recognition Algorithm Based on Convolutional Neural Network and Bayesian Method," in IEEE Access, vol. 12, pp. 188072-188083, 2024, doi: 10.1109/ACCESS.2024.3512671.

[17] J. Gurrola-Ramos, T. Alarcon, O. Dalmau and J. V. Manjón, "MRI Rician Noise Reduction Using Recurrent Convolutional Neural Networks," in IEEE Access, vol. 12, pp. 128272-128284, 2024, doi: 10.1109/ACCESS.2024.3446791.

[18] A. Ielo et al., "A Convolutional Neural Network Model for Classifying Resting Tremor Amplitude in Parkinson's Disease," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 33, pp. 2034-2043, 2025, doi: 10.1109/TNSRE.2025.3574999.

[19] E. S. Rahayu, E. M. Yuniarno, I. K. E. Purnama and M. H. Purnomo, "A Combination Model of Shifting Joint Angle Changes With 3D-Deep Convolutional Neural Network to Recognize Human Activity," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 1078-1089, 2024, doi: 10.1109/TNSRE.2024.3371474.

[20] G. Baldini, "Classification of Fiber Optics Anomalies Using Transforms Ensemble, Adaptive Smoothing Based on the Standardized Variable Distances Learning Algorithm and Convolutional Neural Networks," in IEEE Sensors Letters, vol. 9, no. 5, pp. 1-4, May 2025, Art no. 6005704, doi: 10.1109/LSENS.2025.3560144.