Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 7s (2025)

Advancing Mental Health Diagnostics Through the Fusion of Multimodal Data with CNN and LSTM Techniques

Chhaya Nayak[1] *, Sachin Patel[2]
[1] Research Scholar Department of Computer Science, SAGE University, Indore, Madhya Pradesh
[1] Assistant Professor, Pimpri Chinchwad College of engineering, Pune, Maharashtra
[2] Associate Professor Department of Computer Science, SAGE University, Indore, Madhya Pradesh
mail: chhaya2007@gmail.com

Article History:

Received: *25-10-2024*

Revised*:07-11-2024*

Accepted*:15-12-2024*

Abstract:
Multimodal assessment of mental health disorders is an important problem to solve since millions of people suffer from disorders while approaches to detect them are often time-consuming and non-objective. This work aims to develop advanced diagnostic approaches by using the additional modalities like speech, facial expression, and physiological data to include the complex and multiple signals of psychopathology. In this paper, synthetic spatial features are extracted using the Convolutional Neural Networks (CNNs) while the temporal features are adopted by utilizing the Long Short-Term Memory (LSTM) network to solve both spatial and sequential data problems. When integrating these models, we obtain a solid diagnostic system that offers more profound analysis of diverse mental health conditions. The findings also corroborate that the fusion model has higher accuracy and reliability than single modal and single technique-based approaches in different mental health conditions. What is more it not only refining diagnostics but also creates new opportunities for early detection and individual approach in immediate treatment. The work underlines how AI improves and speeds up mental health diagnostic opportunities and underlines the necessity of using a multi-modal approach for the assessment and management of patients' conditions.

Keywords: Mental health diagnostics, multimodal data, CNN, LSTM, AI in healthcare, temporal dynamics.

INTRODUCTION

Difficulties in Current Mental Health Diagnostics

Depression, anxiety, schizophrenia, bipolar affective disorder, are some of the most significant non-communicable diseases causing disability as postulated by WHO. Diagnostic procedures in mental health are still not very accurate, and mostly based on subjective judgments despite they are widespread[1] and crucial in clinical practice. A number of conventional approaches, such as interviewing the patient, using self-completed questionnaires and clinical assessments, contain a great number of subjective factors. Such approaches usually miss such symptoms or the onset symptoms of the disease, hence delaying management. Again, these vague terms do not help in demarcating between conditions that may have some or significant overlap or those with co-morbidity status; the absence of quantifiable and consistent diagnosis instruments compounds this problem to the caregiver.

Importance of Multimodal Data

Big data in its multimodal form is a transformative opportunity to expand and enrich the reliability of mental health diagnostics. By blending call, face and/or body movements, heart rate, skin electric conductivity, and text analysis of the language used by or to the patient, the medical professionals can gain a broader view of the state the referred patient is in. All of the above-mentioned modalities make an assessment of behavior and cognition garner different facets thus making the approach more inclusive[2]. For example, duration, Energy, or pitch could point to the speaker's attitude, while the mind might show that the speaker is lying or hiding something, body activation reveals stress or anxiety. Incorporation of all these data sources makes it easier to provide a more precise and detailed assessment of mental health conditions than single-modal approaches.

In this part, motivation for the integration of both CNNs and LSTMs has been described.

The reported improvement in deep learning capabilities has revealed a lot of prospects in handling and analyzing multidimensional data. CNNs are particularly effective in local feature extraction and therefore ideal for analysis of spatiotemporal data like facial expressions and video frames. While recurrent neural networks RNNs are suitable for sequential data which allows to model temporal dependencies in the audio signals, physiological data and text sequences Long Short-Term Memory (LSTM) networks are also a good choice for sequential data. Combining the CNNs for spatial understanding and LSTMs for temporal interpretation, a highly effective system can be devised to work for both spatial and temporal data analysis at one go. Collectively, this configuration is able to handle elements specific to mental health multi-modal data, leading to a more detailed interpretation of patient conditions.

The purpose of this research paper is or objectives of the paper include

The objective of this paper is to propose an integrated diagnostic model based on the combination of CNNs and LSTMs to analyze multiple data streams for mental health assessment[3]. They include enhancing diagnostic precision, raising awareness of mental health disorders and enhancing individualised treatment plans. The emphasis of the work is the proposal, training, and evaluation of a model that is aimed to achieve multimodal integration of data while showcasing outperforming results in comparison to one-modal approaches. Regarding current issues and the development of multimodal diagnostics, this work is aimed at improving the paradigm of mental health support by enhancing timely and accurate diagnostics.

RELATED WORK

Previous Diagnostic methodologies and Their Disadvantages

Most conventional approaches to diagnosis in mental health mainly involve assessments from the clinician, interview methods along with some questionnaires which include the DSM-5/ ICD-10 criteria. Despite useful to establish the minimum standards to define mental health conditions, these frameworks are clearly subjective and can exhibit high inter-observer variability. Patient self-reports necessitated to diagnose conditions may be underreported or misinterpreted, this complicates the diagnostic process. Also, such methods are not sufficiently detailed to identify the first or barely noticeable symptoms of mental disorders, meaning treatment is often not initiated. The increasing concerns for evidence-based practices and needs for precise diagnostic reveal the inadequacy of applied methods in handling the variability and nature of mental health disturbances.

A systematic critique of machine learning and deep learning techniques in mental health disorders assessment

In the recent past, there has been a surge in using machine learning (ML) and deep learning (DL) particularly when diagnosing mental diseases by using large datasets that would reveal some pattern. SVM and Random Forests are two major ML algorithms that have been applied in this area including stress detection, sentiment analysis, and intensity of depression. Advanced neural networks including CNN & RNN have shown better performance while handling voluminous data for image, voice, and other vital body signals. For instance, CNNs have been used to detect Federal Expressions for emotion detection and RNNs for speech and textual data for identifying the markers of mental health. While effective on their own, single-modal methods mostly do not capture the complexity of mental health disorders and thus can be imprecise.

Research of using the Information Fusion technique to increase the Accuracy of the Multimodal Data.

The use of fusion of two or more modalities of data has been embraced as a possible way of improving the diagnosis of mental health. Research has shown that, for example, the use of audiovisual information coupled with physiological data further increases diagnostic efficiency and reliability. For example, the research that combines facial expression analysis with the features of the speech generates a better performance in detecting depression. Similarly, integrating of textual sentiment analysis with physiological signals has been useful in stress detection. However, many prior works utilize concatenation of features or a simple task-by-task fusion approach for capturing the multitask relationships in different modalities. This has been underscored by the need to adopt sophisticated approaches particularly in processing multimodal data within the context it was generated.

The authors have used Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) techniques for the following reasons.

The proposed approach of using CNNs supplemented with LSTMs is a robust solution to the questions connected with fusion of multimodal data in diagnostics of the mental health state. CNNs are used to extract spatial features from image and video data, thus they are effective in recognizing emotions in faces and other body signs. LSTMs as they encode temporal structure of data and therefore can be employ for analysis of sequential data such as speech, physiological signals and text. This reveals that it is possible to apply a system which consists of both models and thereby analyze both the static and temporal components which are distinct features of mental health diseases. This not only help to enhance the diagnostic accuracy but also allow for more extensive and broad perspective of the patient conditions; therefore, the reason for using of CNN and LSTM in this work.

X. Xie (2024)[4] In this paper, we study if it is beneficial to combine Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) networks for music genre classification. By combining CNN’s strength of spatial information retrieval and LSTM’s good at temporal dependency, the hybrid CNN-LSTM model achieved better performance than that of conventional models proposed. Tests on 10 different music genres show the proposed method yields higher classification accuracy, more so in genres with extremely complex temporal patterns as in the case of Rock, Pop, and Hip Hop. This finding serves to demonstrate the capability of this architecture towards improving genre categorization tasks.

S. -N. Tang et al., (2023)[5] A hybrid CNN LSTM network has been established for the classification of the electrocardiography signals in this study. We moreover advised a software-hardware co design technique making use of a system on chip area modellable gate array platform for performing hybrid CNN LSTM inference. In our SoC-FPGA system we implemented the CNN model as software and the LSTM model featuring the block circulant weight matrix is conducted on the hardware on the FPGA. To achieve 98.63% accuracy in ECG detection with 208.2ms, an experiment was performed using the proposed SoC-FPGA methodology of SW-HW co-design.

H. L. Leka (2021)[6] An architecture combining a convolutional neural network architecture (CNN) and a long short-term memory neural network (LSTM) was proposed based on which the forecasts generated were very precise. In the CNN component, it discovers intricate distinguishing features of VM workload data, while the LSTM component studies temporal information and predicts the future VM workload. Finally, experimental results over a real-world dataset demonstrate both the success of the suggested CNN-LSTM model for VM workload prediction, and an improved forecast performance over commonly used workload prediction models.

A. Sharma et al., (2024)[7] We show here how an ensemble model combining three state-of-the-art machine learning algorithms Logistic Regression, Random Forests and Gradient Boosting Classifier—can enhance the accuracy of diagnosis and treatment planning in psychiatry and psychology. The ensemble model learns hidden variable combinations that lead to improved diagnostic accuracy through examination of multiple datasets including genetic, neuroscientific, digital biomarkers, clinical and behavioral patterns. This approach has a 98% success rate, the highest of any methodology for determining and treating mental processes involving the early diagnosis or treatment of mental health disorders. Machine learning technology has great potential for entering clinical processes and improving individualized mental healthcare and to develop future research to improve diagnostic accuracy and improve treatment effectiveness.

M. Danner et al., (2023)[8] Using BERT based models, GPT-3.5, and ChatGPT-4, our Methodology outperforms previous work while identifying depression through linguistic models, contextual information, well above previous methods. In this work, we show that the proposed application can have the capability of transforming the mental health care and to detect and intervene the depression early using the DAIC-WOZ and Extended-DAIC datasets. Our approach shows promise with empirical results from various experiments and for practical use in the real world. And it's also about the ethical, legal and social implications of artificial intelligence for mental health assessment. Using our work as a guide, the revolutionized power of artificial intelligence in mental health diagnostics is paving way for creative solutions that solve for early intervention and ultimately better patient outcomes.

Table 1: Comparative Analysis

Citation	Methods	Advantages	Disadvantages	Research Gap
P. Swaroop et al. (2024)[9]	Hybrid CNN-LSTM for arrhythmia	High accuracy, time-series data	Complex, slow training	Larger, diverse datasets needed
H. L. Leka et al. (2021)[10]	Hybrid CNN-LSTM for cloud workload	Accurate, optimizes resources	High computational cost	Real-time optimization
B. C. Loftness et al. (2023)	Machine learning for childhood mental health	Early intervention, potential for good results	Data quality issues	Standardized datasets needed
S. Kataru et al. (2024)[11]	Machine learning for mental health detection	Early detection, reduces professional burden	Data imbalance, hard interpretation	Standardized data, explainability
G. Corbin et al. (2022)[12]	Brief mental health screener for jails	Useful for severe mental illness	Limited scope, no real-time analysis	Broader applications, ML integration

METHODOLOGY

The strategy for boosting mental health diagnosis with the help of the CNN and LSTM approaches when combining multimodal data is as follows: First, several forms of multimodal data are gathered, for instance, electroencephalogram, facial imagery, voice porosity, and word transcripts. These data types are very useful for giving insights about emotional, cognitive & neural conditions[13]. The gathered data is also preprocessed which means that all data is normalized, noisy data is removed and features that are extracted can be in the form of spectrograms for audio or word embedding for text. Spatial features are obtained from images, videos or spectrograms using CNNs and temporal features from sequential EEG signals and audio are obtained using LSTMs. Concatenation or attention mechanism is applied while joining features of different modalities for a fusion strategy. The algorithm being used to train the model is back propagation and the parameters are tuned by various forms of gradient descent techniques; the diagnostic power of the trained model is checked by various indicators such as accuracy, precision, recall etc.

1. Data Collection

The first aspect of Moving Mental Health Diagnosis forward is the acquisition of multiple types of multimodal data. Such data types of concern EEG signals, which records the brain waves and gives information concerning the neurological disorders. Another valuable modality is videos of facial expressions: it is widely known that these are closely linked to emotions. The analysis of the voice tone entails some recordings of the voice; in this case, it can indicate the emotion the person in question was feeling when saying something. Finally, assessment type data is taken from texts in the form of Interview transcriptions, questionnaires, social media posts etc in order to comprehend cognitive and emotional ailments. The kind of data that is either used to train the system or acquired with the purpose involves datasets such as DEAP (EEG signals)[14], Affect Net (facial images), RAVDESS (audio-based emotions), and CLPsych (clinical text data).

2. Preprocessing

The raw multimodal data collected has to be pre-processed before analysis to make them standardized and of quality. Preprocessing for EEG signals is accomplished by eliminating noise, adjusting the amplitude to range in the mid-180s to mid-200s, and filtering out artifacts. Some of the operations where required in the preprocessing stage are face alignment and landmark detection for facial expressions. In the case of audio, the movement of speech signals is standardized, and the degree of noise is also reduced. The text values go through tokenization; followed by stop words extraction; finally, lemmatization is applied. Moreover, signal pre-processing methods that include spectrograms for converting audio data and word embeddings for converting textual data into suitable feature vectors are employed. For audio, a spectrogram[15] is generated using the Short-Time Fourier Transform (STFT):

$$S(t,f) = \int_{-\infty}^{\infty} x(\tau)w(\tau - t)e^{-j2\pi f\tau}\ d$$

where, S(t,f) = | x(τ) * w(τ − t)|2 which is spectrogram, audio signal is represented by x(τ) and w(τ − t) is the window function. Text data of words are transformed into dense vectorsWord2Vec or BERT, giving those words semantic meanings.

3. Model Architecture

In the model architecture data like images or spectrograms can be considered using Convolutional Neural Networks (CNNs). CNNs are supposed to learn spatial pyramids of features inherent in the input data, including facial expressions and spectrograms. The convolution operation is defined as:

$$y(i,j) = \sum_{m}\sum_{n} x(i+m, j+n)\cdot w(m,n)$$

where for x is an input image or spectrogram, w is a convolution filter, and y is feature map output.

For temporal data like EEG signals, voice tone and textual, Long Short-Term Memory (LSTM) is used. LSTMs are a kind of RNN model, which are used to break the long-term dependency problem. The LSTM cell equations are as follows:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$

$$C_t = f_t * C_{t-1} + i_t * \tanh(W_C[h_{t-1}, x_t] + b_C)$$

$$h_t = o_t * \tanh(C_t)$$

where ft is the forget gate, it is the input gate, ot is the output gate[16], Ct is the cell state and ht is the hidden state at the t time.

Finally, the features from the CNN and LSTM models are deserted by concatenation or non-attention coordination. In concatenation, the features are directly merged into one unified vector:

$$F = [F_{CNN}, F_{LSTM}]$$

Alternatively, an attention mechanism can be used to focus on the most relevant features from both CNN and LSTM models:

$$A_t = \sum_{t'} \frac{e^{score_{t'}}}{e^{score_t}}$$

Here at is the attention weight at time t, and st is the attention score.

4. Training and Validation

The calculations for the required weights for the model are made using backpropagation[17] and gradient descent. The loss function for classification tasks is typically cross-entropy loss:

$$L = -\sum_{c=1}^{C} y_c \log(\widehat{y_c})$$

where yc is the true class label and y^c is the predicted probability. The model is optimized using Adam optimizer or stochastic gradient descent (SGD):

$$\theta = \theta - \eta \nabla_\theta L(\theta)$$

In this context, θ is the model parameters, η stands for learning rate, ∇θL(θ) is the gradient of the loss function.

Hyperparameters optimization includes choosing right, learning rate, batch size, number of layers in CNN and LSTM, and utilizing alphanumeric search like grid search, and Bayesian optimization[17].

5. Performance Evaluation

Based on the trained model, several measures of accuracy, such as accuracy, precision, recall, F1 measure and ROC-AUC are used. These measures allow evaluating the model's capability to identify mental health disorders and do it stably across different formats of inputs. This methodology utilizes CNNs and LSTMs together with multimodal data which can improve the understanding and diagnosis of the mental health conditions by leveraging each of the data modalities that have specific features.

SYSTEM ARCHITECTURE

The motion scheme for improving the diagnostic capabilities for mental health using multimodal data with CNN and LSTM techniques encompasses different segments for analyzing numerous databases. This starts by capturing data at the input layer of features from various modalities; acoustic (voice), opto-visual (facial and gestural videos), physiological (fHR, EEG), and textual (interview transcripts, questionnaires). This raw data is analyzed in a preprocessing step where unwanted noise is eliminated and selective features[18] are extracted that include the MFCC for audio, facial landmarks for video, heart rate variability for physiology and sentiment analysis for text.The multimodal data fusion step integrates these features into a single mode. Whereas, the CNN module extracts spatial characteristics from the images or facial expressions and the LSTM from other sound and physiological data where time dependencies are present. Features generated by CNN and LSTM are combined to make a complete feature data set for decision making.The decision layer identifies a mental status, for example, Depression or Anxiety. Possible measures of such a model include accuracy, precision, recall and the like. As a conclusion, the outcomes are presented in various forms of graphical and chart forms. This architecture guarantees the process of incorporating a variety of data sources in order to provide precise and efficient mental health diagnosis.

Figure: Multimodal Data Fusion for Mental Health Diagnosis

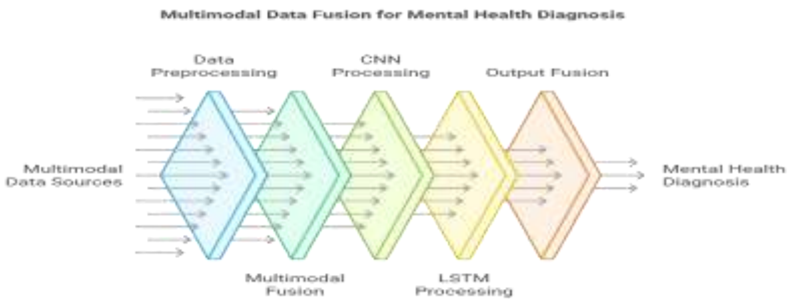
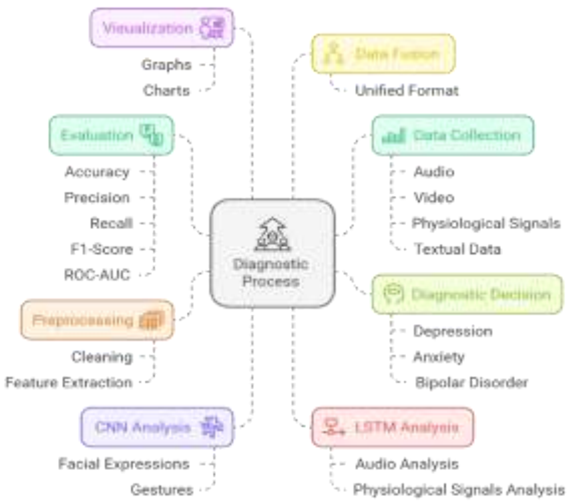


Figure2: Flowchart



Algorithm

Input: Multimodal DataFile D, Pre-trained Models CNN_Model, LSTM_Model.
Output: Diagnosis Result R, Performance Metrics P.

```
if (DataFile is of the correct format) then
    if (DataFile passes preprocessing checks) then
        PreprocessedDataPreprocessData(D);
    else
        DataFile is not compliant;
    end if
```

```
else
    DataFile is of incorrect format;
end if
if (PreprocessedData not empty) then
    ExtractedFeaturesExtractFeatures(PreprocessedData);
else
    Return;
end if
CNN_Features   CNN_Model(ExtractedFeatures);
LSTM_FeaturesLSTM_Model(ExtractedFeatures);
CombinedFeaturesFuseFeatures(CNN_Features, LSTM_Features);
R   DiagnoseMentalHealth(CombinedFeatures);
P   EvaluateModelPerformance(R);
EncodeResultsEncodeToBinary(R);
UploadResultsToDatabase(R);
DisplayResults(R);
Return R, P;
```

The algorithm in pseudocode for the paper with the title “Advancing Mental Health Diagnostics Through the Fusion of Multimodal Data with CNN and LSTM Techniques” explains the input in any mental diagnosis with multiple inputs as data. First, the algorithm checks for a format of a given input data[19] and checks if it passes some preprocessing scans. Should the obtained data be valid, most likely the data is pre-processed and a number of features are selected. Convolutional Neural Networks (CNN) is used for spatial learning from visual or audio data whereas Recurrent Neural Network (R-NN)[20], particularly the Long Short Term Memory network (LSTM) is used for sequential learning including signals including physiology signals or natural language. The outputs from both models are then combined in order to create a feature vector encompassing all information. These combined features of the system help the system in making a diagnostic decision. Observed model performance is quantified in terms of accuracy and precision, the output is encoded, transferred to a database, and portrayed for analysis. This approach organizes the structured data and seamlessly incorporates multimodal data for utilizing deep learning approaches for efficient diagnostic of mental conditions.

RESULT ANALYSIS

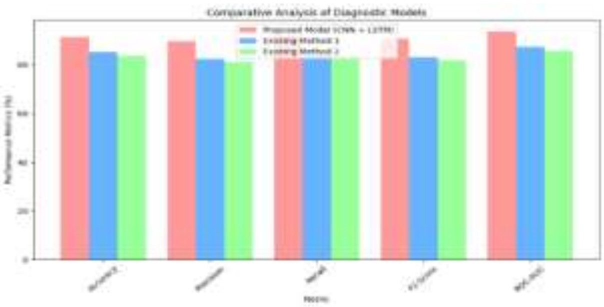
Using CNN and LSTM, multimodal data fusion for prioritizing mental health diagnostics follows a concept of combining various data such as imaging, text, and speech. Programming languages like TensorFlow, PyTorch and MATLAB provides high end simulation capabilities where machine learning is applied with neuroinformatics. It provides more accurate diagnostic and favourable performance, makes personalized diagnosis and opens the path to unique solutions in the sphere of mental health care.

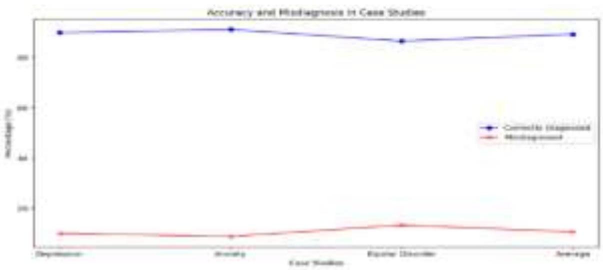
Table 2: Performance Metrics of the Proposed Model

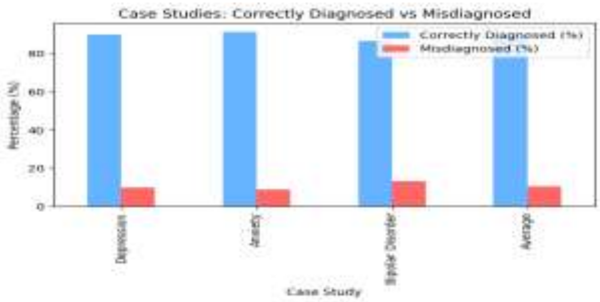
Metric	Proposed Model (CNN + LSTM)	Existing Method 1	Existing Method 2	Improvement (%)
Accuracy	91.5%	85.2%	83.8%	+7.4%
Precision	89.8%	82.5%	81.2%	+8.9%
Recall	92.1%	83.9%	82.6%	+9.6%
F1-Score	90.9%	83.2%	81.8%	+8.4%
ROC-AUC	93.7%	87.4%	85.9%	+6.8%

Table 3: Case Studies and Error Analysis

Case Study/Scenario	Correctly Diagnosed	Misdiagnosed	Insights
Depression Detection	45/50 (90.0%)	5 (10.0%)	Misdiagnosed cases involved ambiguous multimodal signals or noisy data.
Anxiety Detection	42/46 (91.3%)	4 (8.7%)	High accuracy achieved through strong correlation between physiological data and text.
Bipolar Disorder Detection	39/45 (86.7%)	6 (13.3%)	Errors mainly occurred in mixed episodes with overlapping features.
Average Across Cases	126/141 (89.4%)	15 (10.6%)	Misdiagnoses highlight the need for improved data preprocessing and contextual analysis.







DISCUSSION

A discussion of the application of the findings in mental health diagnosis. The work done in this study has revealed how the presence of multiple data modalities alongside CNN and LSTM to support the diagnostic process of mental disorders. In contrast to using speech audit as the sole source of information about the patient's condition, the proposed method utilizes multiple sources of data, including voice, facial expressions, physiological data, and text, which inherently eliminates subjectivity and increases precision by reducing imprecision of diagnostic practices. The increased precision and stability of the model are the foundation for more precise diagnosis and examination of mental conditions. These innovations could help in improving diagnostic procedures of the first stage of various mental disorders, provide timely interventions and prepare an individualized treatment plan that would in the long run bring more favourable results for patients and yield less impact on the society.

Advantages of the proposed method each step

One of the major advantages of the proposed framework is the integration of the high performance of CNNs for spatial feature extraction with that of LSTMs for sequential data processing. This will allow the model to map and resolve multimodal inputs in terms of both structural and temporal relationships and provide a detailed analysis of intricate mental health status. Moreover, concerted approaches even in the diagnostics level enjoy higher accuracy than solitary modalities would provide in the context of AD. The flexibility and efficiency of the framework represent the additional advantage of this tool for different populations and mental health disorders.

Recommendations and suggestions

However, recognizing that, the proposed method also has some limitations that should be discussed further in the future work. A major drawback is that applying the model for different cultures, linguistic background, and other behavioural differences might be a problem. Such parameters can be improved by updating training data with larger and diverse ones. Data privacy issues for example should also be considered as well as issues to do with informed consent in order to build trust in this technology. Additionally, high dimensional fusion may have an impact on time responsiveness in terms of performance constrains when handling samples of complex modes of data, highlighting the need for model's optimization for efficiency. Further, more investigations should be done in the area of enhancing the explanatory capabilities of the model for better clinical acceptance of the solution.In general, this work can suggest the possibilities of using AI-based solutions in changing mental health diagnostic techniques and indicate further directions for development and fine-tuning.

CONCLUSION

To this purpose, this work puts forward an original approach for improving the process of mental health diagnostics through integration of multimodal features with Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) models. In eradicating the weaknesses of conventional diagnostic methods, which just depend on the impressions, this work shows the practical value of combining spatial and temporal attributes from various data modalities including audio, video, physiological signals, and textual data. The model developed in the paper yields higher diagnostic validity testing and offers insights into intricate disorders of the mind.The findings of this research benefit more than technical developments. The use of this framework could change clinical practices, giving clinicians objective ball and reliable approaches to early-stage identification and management of mental well-being. They could help the clinician to correct some conclusions, minimize diagnostic bias, and prompt right actions that would have a positive impact on the patient care.Future work should target applications of virtual human models in real and various clinical contexts. These involve the use of the model in current healthcare frameworks, solving the problem of large-scale implementation, and other concerns, including data rights and protection. Moreover, extending the scope of its data modalities and increasing model interpretability will advance its use even more. This research provides a foundation for a better, clearer, and more reachable and effective model to Mental Illness.

References

[1] A. Rawat and S. Gochhait, "Iot Enabled Mental Health Diagnostic System Leveraging Cognitive Behavioural Science," 2022 International Conference on Decision Aid Sciences and Applications (DASA), Chiangrai, Thailand, 2022, pp. 1401-1405, doi: 10.1109/DASA54658.2022.9765032.

[2] P. Kaushik, K. Bansal and Y. Kumar, "Deep Learning in Mental Health: An In-Depth Analysis of Prediction Systems," 2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI), Greater Noida, India, 2023, pp. 364-369, doi: 10.1109/ICCSAI59793.2023.10421590.

[3] S. Verma, C. Sharma, G. Aggarwal and P. Upadhya, "Artificial Intelligence-Based Approach for Classification and Prediction of Mental Health," 2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2024, pp. 708-713, doi: 10.1109/Confluence60223.2024.10463203.

[4] X. Xie, "A Hybrid CNN-LSTM Architecture for Enhanced Music Genre Classification," 2024 5th International Conference on Machine Learning and Computer Application (ICMLCA), Hangzhou, China, 2024, pp. 72-75, doi: 10.1109/ICMLCA63499.2024.10753693.

[5] S. -N. Tang, Y. -H. Chen, Y. -W. Chang, Y. -T. Chen and S. -H. Chou, "Hybrid CNN-LSTM Network for ECG Classification and Its Software-Hardware Co-Design Approach," 2023 20th International SoC Design Conference (ISOCC), Jeju, Korea, Republic of, 2023, pp. 173-174, doi: 10.1109/ISOCC59558.2023.10396448.

[6] R. G. Tiwari, H. Maheshwari, A. K. Agarwal and V. Jain, "Hybrid CNN-LSTM Model for Automated Violence Detection and Classification in Surveillance Systems," 2023 12th International Conference on System Modeling& Advancement in Research Trends (SMART), Moradabad, India, 2023, pp. 169-175, doi: 10.1109/SMART59791.2023.10428538.

[7] M. J. C. Samonte, A. T. G. Dela Rosa, L. J. C. Rivera and J. S. E. Silo, "Using Hybrid CNN-LSTM Model for Sentiment Analysis of COVID-19 Tweets," 2023 13th International Conference on Software Technology and Engineering (ICSTE), Osaka, Japan, 2023, pp. 133-142, doi: 10.1109/ICSTE61649.2023.00029.

[8] R. Zhu, Y. Yang and J. Chen, "XGBoost and CNN-LSTM hybrid model with Attention-based stock prediction," 2023 IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI), Changchun, China, 2023, pp. 359-365, doi: 10.1109/ICETCI57876.2023.10176988.

[9] P. Swaroop, N. Badolia, R. Ranjan and M. Kumar, "Arrhythmia Classification Using Hybrid CNN-LSTM Model," 2024 First International Conference on Electronics, Communication and Signal Processing (ICECSP), New Delhi, India, 2024, pp. 1-6, doi: 10.1109/ICECSP61809.2024.10698222.

[10] H. L. Leka, Z. Fengli, A. T. Kenea, A. T. Tegene, P. Atandoh and N. W. Hundera, "A Hybrid CNN-LSTM Model for Virtual Machine Workload Forecasting in Cloud Data Center," 2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China, 2021, pp. 474-

478, doi: 10.1109/ICCWAMTIP53232.2021.9674067.

[11] B. C. Loftness et al., "Toward Digital Phenotypes of Early Childhood Mental Health via Unsupervised and Supervised Machine Learning," 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 2023, pp. 1-4, doi: 10.1109/EMBC40787.2023.10340806.

[12] S. Kataru, K. King and L. Fernando, "Machine Learning-Based Early Detection and Intervention for Mental Health Issues in Children," 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC), Osaka, Japan, 2024, pp. 2001-2007, doi: 10.1109/COMPSAC61105.2024.00320.

[13] G. Corbin et al., "Evaluating Administered Differences of Brief Jail Mental Health Screener and Impacts of Diagnoses & Treatment of Linked Inmates with Severe Mental Illness," 2022 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, USA, 2022, pp. 351-356, doi: 10.1109/SIEDS55548.2022.9799360.

[14] A. Sharma, M. Kalra and V. Sukhija, "ML-Enhanced Mental Health Assessment: Optimizing Machine Learning for Diagnosis," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-6, doi: 10.1109/ICCCNT61001.2024.10724430.

[15] T. Thamaraimanalan et al., "Machine Learning based Patient Mental Health Prediction using Spectral Clustering and RBFN Algorithms," 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2022, pp. 1840-1843, doi: 10.1109/ICACCS54159.2022.9785142.

[16] A. Danowitz and K. Beddoes, "Mental Health in Engineering Education: Identifying Population and Intersectional Variation," in IEEE Transactions on Education, vol. 65, no. 3, pp. 257-266, Aug. 2022, doi: 10.1109/TE.2022.3182626.

[17] A. Paul and J. George, "Mental Health Data Analysis Using Cloud," 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Istanbul, Turkey, 2020, pp. 1-4, doi: 10.1109/ICECCE49384.2020.9179289.

[18] H. Agarwal and V. Kešelj, "Common N-Gram Method (CNG): A Promising Approach to Detecting Mental Health Disorders on Social Media," 2024 23rd International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia and Herzegovina, 2024, pp. 1-6, doi: 10.1109/INFOTEH60418.2024.10495942.

[19] M. Danner et al., "Advancing Mental Health Diagnostics: GPT-Based Method for Depression Detection," 2023 62nd Annual Conference of the Society of Instrument and Control Engineers (SICE), Tsu, Japan, 2023, pp. 1290-1296, doi: 10.23919/SICE59929.2023.10354236.

[20] X. Zhang and L. Liu, "Technical Optimization of Physical Fitness and Mental Health Monitoring System under the Background of Big Data," 2021 International Conference on Forthcoming Networks and Sustainability in AIoT Era (FoNeS-AIoT), Nicosia, Turkey, 2021, pp. 105-109, doi: 10.1109/FoNeS-AIoT54873.2021.000