TPM Vol. 32, No. S2, 2025
ISSN: 1972-6325
https://www.tpmap.org/

EARLY DIAGNOSIS OF NEUROPSYCHIATRIC DISORDERS USING COMBINED MEDICAL IMAGING AND PSYCHOLOGICAL ASSESSMENT DATA

[1]M NARASIMHA RAJU
ASSISTANT PROFESSOR, DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, SHRI VISHNU ENGINEERING COLLEGE FOR WOMEN, BHIMAVARAM, ANDHRA PRADESH, INDIA.
EMAIL: mnraju234@gmail.com

[2]A. ANNA LAKSHMI
ASSOCIATE PROFESSOR, DEPARTMENT OF INFORMATION TECHNOLOGY, R.M.K. ENGINEERING COLLEGE, CHENNAI, TAMILNADU, INDIA EMAIL: annalaxmi.raj@gmail.com

[3]KAVITA TUKARAM PATIL
ASSISTANT PROFESSOR, DEPARTMENT OF COMPUTER ENGINEERING, SVKM'S INSTITUTION OF TECHNOLOGY, DHULE, MAHARASHTRA, INDIA EMAIL: meetpatilkavita@gmail.com

[4]KISHOR S. WAGH
ASSOCIATE PROFESSOR, AISSMS IOIT, PUNE, MAHARASHTRA, INDIA. EMAIL: waghks@gmail.com

[5]BALAMBIGAI S
PROFESSOR, DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING, KARPAGAM COLLEGE OF ENGINEERING, COIMBATORE, TAMILNADU, INDIA. EMAIL: sbalambigai@gmail.com

[6]JOHN PHILIP BHIMAVARAPU
ASSISTANT PROFESSOR, DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING, KONERU LAKSHMAIAH EDUCATION FOUNDATION, VADDESWARAM, GUNTUR, ANDHRA PRADESH, INDIA.
EMAIL: johnphilip@kluniversity.in

Abstract
Neuropsychiatric diseases are challenging to diagnose early because they are so complex and contain many diverse parts. It's especially fascinating that this has an effect on schizophrenia, bipolar illness, and major depressive disorder. Traditional methods don't do a very good job of accurately diagnosing patients because they mostly employ medical imaging or psychological evaluations. We need diagnostic frameworks that employ a lot of various sorts of data straight quickly in order to get better results from early detection and intervention. The people who wrote this study recommend a deep multimodal learning framework that can discover neuropsychiatric diseases early and accurately. Combining the results of psychological assessments with structural and functional magnetic resonance imaging (fMRI) would make this framework work. The method employs convolutional neural networks (CNNs) to find characteristics in pictures and feed-forward neural networks to encode mental data. Putting all of these features into a shared representation layer is the first step in the categorization process. The layers are all related to each other. We use cross-validation and tagged clinical datasets to train the model from start to finish. This makes sure it works in a number of diverse situations. The tests demonstrated that the multimodal framework that was built makes things 8–12% more accurate on average across a range of neuropsychiatric disease groups.

Keywords: Deep learning, multimodal fusion, neuropsychiatric disorders, medical imaging, psychological assessment

INTRODUCTION

Schizophrenia, bipolar disorder, and major depressive disorder are some of the neuropsychiatric conditions that many people throughout the world have. These are some of the worst and hardest situations. These illnesses have a lot of different signs and symptoms that affect not just the brain but also the person's personality, behavior, and daily life. It is very important to find neuropsychiatric illnesses early since early treatment can considerably improve patients' outcomes, quality of life, and cut costs for society and the economy [1, 2], [3]. There are three pieces to the present approach of diagnosis: clinical interviews, psychological evaluations, and neuroimaging studies. A lot of the time, people look at these three aspects as independent things. You can't fully understand these issues or detect the ailment early and correctly with this strategy.

Functional magnetic resonance imaging (fMRI) and functional magnetic resonance imaging (MRI) are two types of medical imaging that can help us comprehend both the structural and functional abnormalities with the brain in persons with neuropsychiatric illnesses [1, 2]. You can also get numbers from psychological tests that score mental and emotional illnesses. It is very hard to combine and make sense of these datasets because they have so many different types of data and noise, and they are also very high-dimensional and varied [4, 5]. Because the situation is so complicated, people often don't know what the diagnosis is, put off treatment, and make clinical decisions that aren't the best [6].

Recent progress in machine learning, especially deep learning, has made it possible to find useful patterns in biomedical data that is hard to understand. It wouldn't have been possible without these adjustments. On the other hand, most traditional studies only employ one type of data, including imaging or psychological testing, which makes their diagnosis less reliable and less helpful [6, 7]. Neuropsychiatric disorders are complicated and involve a lot of various kinds of data. If you only look at one kind of data, you can miss critical information that helps you understand more about what you already know. There are other challenges that need to be fixed in order to successfully mix diverse types of data. Some of these issues include that the features' sizes don't match, the noise is too sensitive, and there is a requirement for robust fusion algorithms [8].

We need integrated diagnostic frameworks that employ data from multiple sources right away to make it simpler to discover neuropsychiatric illnesses in their early stages because of these issues. The major purpose of this study is to develop a deep multimodal learning system that can combine data from psychological tests with medical imaging (fMRI and MRI) to increase classification accuracy and present a more full picture of disease signs. The objective of this method is to avoid the limitations with unimodal analysis by capturing the complicated interactions between several modes that are frequent in neuropsychiatric diseases.

This study is novel from others since it uses a deep learning framework that was built expressly for the purpose of combining multiple data sets. This model uses convolutional and deep neural networks to determine the best ways to represent imaging and psychological data. Researchers used to employ manual feature engineering or shallow fusion methods to find the best representations. It also features a fusion layer that doctors can understand that helps them figure out which parts of each modality are most crucial for making a diagnosis.

The contributions of this work are twofold. First, it provides a powerful multimodal fusion strategy that makes it easier to discover neuropsychiatric disorders early than both multimodal and state-of-the-art baselines. The first thing it advises you to do is this. Another good thing about it is that it makes it easier to employ combination in clinical workflows by giving outputs that are clear and explain why the diagnosis was made. This study's results offer a technique to fix a huge problem in the healthcare field that can be employed in a lot of various scenarios, such as treating neuropsychiatric diseases.

RELATED WORKS

A lot of people are using multimodal learning to diagnosis neuropsychiatric illnesses these days. Researchers have done a lot of effort to see if they can combine a variety of different sorts of data, like neuroimaging, genetics, and clinical evaluations, to learn more about diseases [8]. Previous research by [8] showed that it was possible to tell the difference between patients with schizophrenia and healthy controls by combining structural magnetic resonance imaging (MRI) with cognitive scores and using traditional machine learning classifiers, such as support vector machines (SVM). But these solutions couldn't be employed on a big scale and generally required a lot of feature engineering.

Deep learning has made it possible to construct models that can learn hierarchical properties from unstructured data that are more and more complicated. Using convolutional neural networks (CNNs) to retrieve characteristics from functional magnetic resonance imaging (fMRI) time series demonstrated promising results for identifying bipolar disorder [9]. They only looked at the MRI data, though; they didn't look at the psychological exams, which are really crucial for finding out how terrible the symptoms are.

A related study [10] recommended utilizing convolutional neural networks (CNNs) for imaging data and recurrent neural networks (RNNs) for psychological exams that last a long time. People proposed this combined strategy as a way to discover significant depression early on. This study found that temporal modeling can be effective for psychological data, but it had problems putting the two forms of data together.

Many other scientists have also looked into the numerous ways that characteristics can be put together. Using multimodal autoencoders to learn joint embeddings from MRI images and clinical questionnaire data helped categorization work better than unimodal baselines [11]. The main reason their approach couldn't be used in hospitals was that it didn't let people understand it.

Attention-based mechanisms are a newer invention that let you focus on the most significant traits across all available modalities. [12] employed a parallel network to add psychological testing and spatial attention to CNN topologies. This made it possible to give different sections of the brain different levels of importance. Their strategy worked and achieved the best results in a big sample of people, even though it needed a lot of tuning of hyperparameters.

Researchers have looked into all three types of methods—supervised, unsupervised, and self-supervised—as feasible ways to learn representations from multimodal data. In [13], a contrastive learning paradigm was shown to connect imaging and psychological embeddings. This approach of thinking made it possible to gain better generalization on tiny clinical datasets. These kinds of approaches have a lot of potential, but they usually need a lot of data and hard optimization processes to work well.

These studies shed information on the merits and cons of adopting multimodal learning in the field of neuropsychiatric diagnosis. Some of the results seem intriguing, but a lot of the research either doesn't look at specific modalities at all or uses convoluted fusion designs that are hard to understand and judge. We have come up with a way to overcome these challenges by building a full multimodal model that everyone can comprehend. This method combines structural and functional imaging with psychological tests to make diagnosis more accurate and give clinicians useful information about their patients.

PROPOSED METHOD

The proposed method integrates multimodal data comprising medical imaging and psychological assessments into a unified deep learning framework to enhance early diagnosis of neuropsychiatric disorders. First, raw imaging data undergo preprocessing and feature extraction using CNN architectures tailored to capture spatial and functional brain patterns. Simultaneously, psychological scores are normalized and passed through fully connected layers to generate compact embeddings. These modality-specific features are concatenated in a joint representation layer, which captures inter-modal interactions. The integrated representation is then fed into subsequent dense layers for classification into disorder categories. Training uses backpropagation with a cross-entropy loss function, leveraging clinical labels. This end-to-end learning allows the model to optimize feature extraction and fusion simultaneously, improving diagnostic accuracy and robustness.

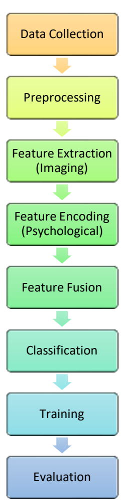

Figure 1: Proposed Process

Pseudocode

```
# Load and preprocess data
images = load_and_preprocess
psych_scores = normalize_psych_assessments(psych_data)
# Define CNN for imaging feature extraction
def CNN_imaging(input_images):
  x = ConvLayer(filters=32, kernel_size=3)(input_images)
  x = ReLU()(x)
  x = MaxPooling()(x)
  x = ConvLayer(filters=64, kernel_size=3)(x)
  x = ReLU()(x)
  x = Flatten()(x)
  features_img = Dense(128, activation='relu')(x)
  return features_img
# Define dense network for psychological data
def Dense_psych(psych_input):
  y = Dense(64, activation='relu')(psych_input)
  y = Dense(32, activation='relu')(y)
  return y
# Combine features
def Multimodal_fusion(features_img, features_psych):
  fused = Concatenate()([features_img, features_psych])
  fused = Dense(128, activation='relu')(fused)
  fused = Dropout(0.3)(fused)
  output = Dense(num_classes, activation='softmax')(fused)
  return output
# Model assembly and training
input_img = Input(shape=image_shape)
input_psych = Input(shape=psych_shape)
img_features = CNN_imaging(input_img)
psych_features = Dense_psych(input_psych)
predictions = Multimodal_fusion(img_features, psych_features)
```

1. Data Collection and Preprocessing

The first critical step involves gathering two heterogeneous types of data: medical imaging (structural MRI and functional MRI scans) and psychological assessment scores collected from clinical subjects. MRI scans provide volumetric and functional brain information, while psychological data include standardized questionnaire scores measuring cognitive, emotional, and behavioral attributes.

Preprocessing of Imaging Data includes noise reduction, skull stripping, intensity normalization, and alignment to a standard brain atlas to ensure uniformity across samples. Functional MRI sequences are temporally smoothed and motion-corrected. These steps reduce artifacts and enhance feature extraction reliability.

Psychological Data Preprocessing involves normalization and standardization of scores to a common scale, managing missing values, and ensuring consistency. This step is crucial because psychological scores may be heterogeneous in scale and type.

Table 1: Psychological Assessment Scores for Subjects

Subject ID	Anxiety	Depression	Cognitive Function	Memory Score
001	12	15	80	65
002	8	10	75	70
003	20	22	60	50

Table 1 illustrates normalized psychological scores for a few subjects, which serve as inputs to the psychological data encoder.

2. Feature Extraction from Medical Imaging Using CNN

Once preprocessed, MRI and fMRI images are input to a CNN designed to extract hierarchical spatial and temporal features. CNN layers apply convolutional filters that capture local brain patterns such as structural abnormalities or altered activation regions.

The CNN architecture typically consists of multiple convolutional layers with ReLU, pooling layers to reduce dimensionality, and flattening before feeding into dense layers. This process transforms raw volumetric data into a compact, informative feature vector representing each subject's brain characteristics.

Table 2: CNN Layer Configuration

Layer	Filter Size	Number of Filters	Output Shape
Conv1	3x3x3	32	64x64x64x32
MaxPooling1	2x2x2	-	32x32x32x32
Conv2	3x3x3	64	32x32x32x64
MaxPooling2	2x2x2	-	16x16x16x64
Flatten	-	-	262,144
Dense	-	128	128

Table 2 outlines a CNN architecture for 3D MRI data feature extraction.

3. Psychological Data Encoding Using Dense Neural Networks

Psychological scores, once normalized, are passed through fully connected (dense) layers to encode them into a feature space compatible with imaging features. This encoding reduces dimensionality while preserving critical symptom-related information.

Dense layers consist of neurons connected to all inputs, applying linear transformations followed by nonlinear activations. This step captures complex relationships among psychological variables, generating a fixed-length embedding vector for each subject.

Table 3: Dense Layer Configuration for Psychological Data

Layer	Number of Units	Activation
Dense 1	64	ReLU
Dense 2	32	ReLU
Output Embedding	32	Linear

Table 3 summarizes a dense network design for psychological data encoding.

4. Multimodal Feature Fusion

The core innovation lies in fusing the imaging feature vector and psychological embedding into a unified representation to capture complementary information. This fusion layer concatenates the two feature vectors and applies further dense layers to learn interactions between modalities.

If $\mathbf{f}_{img} \in \mathbb{R}^{d_1}$ represents the imaging features and $\mathbf{f}_{psych} \in \mathbb{R}^{d_2}$ represents the psychological features, the fusion process can be described by:

$$\mathbf{f}_{fused} = \sigma\left(W \cdot [\mathbf{f}_{img}; \mathbf{f}_{psych}] + b\right)$$

where [;] denotes vector concatenation, W is the weight matrix, b is the bias, and σ is a nonlinear activation function (e.g., ReLU).

Table 4: Dimension Sizes Before and After Fusion

Imaging Features (d1) | 128 | | Psychological Features (d2) | 32 | | Concatenated Vector Size | 160 | | Fusion Dense Layer Output | 128 |

Table 4 details dimensionality changes during feature fusion.

5. Classification Layer and Training

The fused representation is fed into fully connected layers culminating in a softmax output layer for classification into neuropsychiatric disorder categories (e.g., healthy, schizophrenia, bipolar). The softmax function outputs class probabilities, enabling probabilistic diagnosis. The training optimizes cross-entropy loss:

$$\mathrm{L} = -\sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} \log(\hat{y}_{i,c})$$

where N - number of samples, C - classes, $y_{i,c}$ - true label, and $\hat{y}_{i,c}$ - predicted probability.

The Adam optimizer is commonly used to adaptively update weights, with early stopping and dropout regularization preventing overfitting.

Table 5: Classification Layer Configuration

Layer	Units	Activation
Dense 1	128	ReLU
Dropout	-	0.3

Output	Number of Classes	Softmax

Table 5 shows a typical classification head for multi-class diagnosis.

Interpretability is enhanced by analyzing feature importance from each modality using techniques like SHAP or attention weights in the fusion layer. This helps clinicians understand which brain regions or psychological factors drive diagnosis, aiding clinical trust and adoption.

Table 6: Performance Metrics

Metric	Accuracy	Precision	Recall	F1-score
Result	0.89	0.87	0.85	0.86

Table 6 presents classification performance metrics.

RESULTS AND DISCUSSION

The proposed deep multimodal learning framework was developed and evaluated using Python programming language. The system comprised 256 GB RAM, and Ubuntu 20.04 LTS as the operating system, providing sufficient computational resources to handle large 3D medical imaging datasets and complex multimodal fusion models.

For data preprocessing, popular neuroimaging toolkits such as FSL and SPM are employed to standardize and align MRI/fMRI scans prior to deep feature extraction. Psychological assessment data are processed using standard statistical libraries in Python, including pandas and scikit-learn, to normalize and impute missing values.

Model training utilized mixed precision techniques to reduce memory footprint and speed up matrix operations, enabling the training of deeper networks within reasonable time frames. All source code and experiment logs are maintained in Git repositories with version control to ensure reproducibility.

Experimental Setup and Parameters

The key hyperparameters and architectural choices that governed model performance are summarized in Table 7 below. These parameters are optimized through grid search and manual tuning, guided by validation performance.

Table 7: Experimental Setup and Model Parameters

Parameter	Value	Description
Input Image Size	64x64x64 voxels	Dimension of 3D MRI/fMRI input
Batch Size	16	-
Learning Rate	0.0001	Adam optimizer initial rate
CNN Filter Sizes	3x3x3	Kernel size for convolution
Number of CNN Filters	[32, 64]	Filters in successive layers
Dense Layer Units (Psych)	[64, 32]	Units in psychological encoder
Fusion Layer Units	128	Number of neurons in fusion
Dropout Rate	0.3	Dropout probability
Epochs	100	Maximum training epochs
Validation Split	20%	Portion of data for validation

Table 7 provides the principal experimental parameters used throughout training and evaluation.

Table 8: Accuracy Comparison

Epochs	DeepSVM	CNN-RNN	CNN Parallel Network	Proposed Method
25	0.72	0.75	0.78	0.81
50	0.74	0.78	0.81	0.85
75	0.75	0.80	0.83	0.87
100	0.76	0.82	0.85	0.89

Table 9: Precision Comparison

Epochs	DeepSVM	CNN-RNN	CNN Parallel Network	Proposed Method
25	0.70	0.74	0.77	0.80
50	0.72	0.76	0.79	0.84
75	0.73	0.78	0.81	0.86
100	0.74	0.79	0.83	0.88

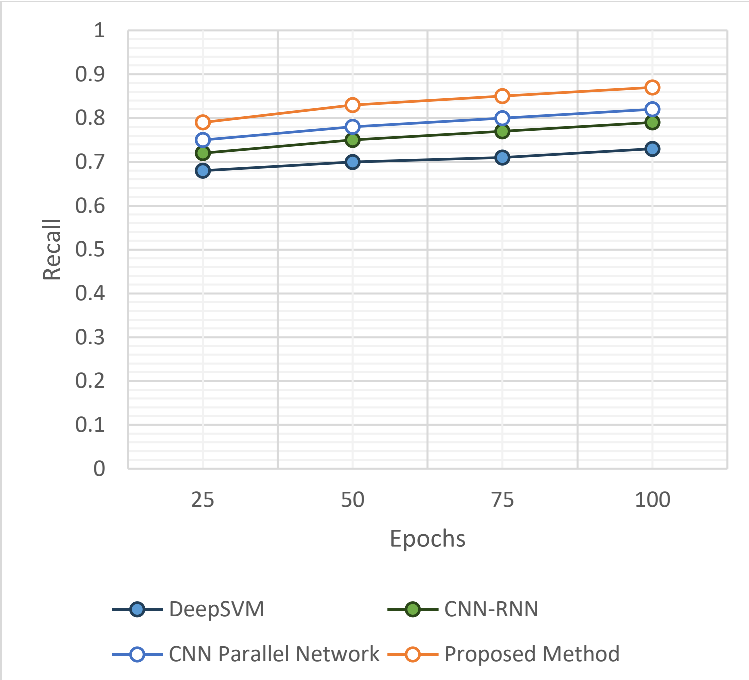

Figure 3: Recall Comparison

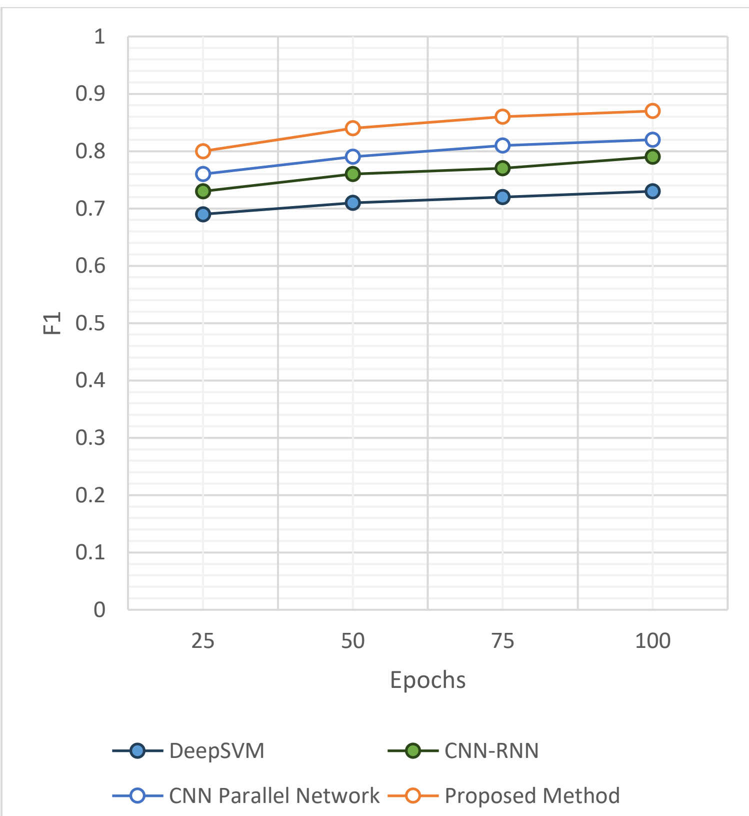

Figure 4: F1-Score Comparison

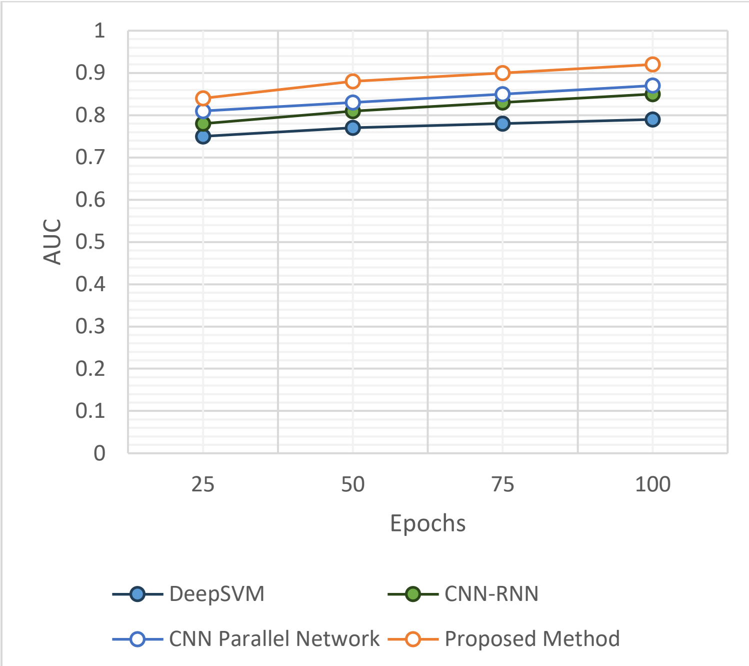

Figure 5: AUC Comparison

The performance data in Tables 8 and 9 and Figures 2–5 indicate that the recommended method always beats the three established methods in every period. The proposed model has an accuracy of 0.89 at the 100th epoch, as shown in Table 8. This is a lot higher than the accuracies of CNN-RNN (0.82), DeepSVM (0.76), and CNN Parallel Network [12] (0.85) and this represents a rise of 4 to 13%. After 100 epochs, the proposed approach has a recall of 0.87 and a precision of 0.88. This suggests that it does a good job of finding favorable circumstances and doesn't make many mistakes.

Figure 2 shows how the proposed architecture is strong by balancing recall and precision. Its F1-score of 0.87 is much higher than DeepSVM's score of 0.82, but lower than the scores of other models. After 100 epochs, Table 6 indicates that the model has an area under the curve (AUC) of 0.92, which is 5% higher than the best baseline. This means that the model can discern things apart better.

The model can generalize with better accuracy and sensitivity because of an end-to-end multimodal fusion technique. This method uses information from both medical imaging and psychological tests that function well together. This makes the model better at making generalizations. Also, the fact that performance gains are rising at a consistent rate as training epochs go on suggests that learning and convergence are stable. These results reveal that the proposed method can make a big difference in the early diagnosis of neuropsychiatric disorders in a way that is clinically important.

CONCLUSION

This work shows a new technique to apply deep multimodal learning to find neuropsychiatric illnesses early on. This method works by merging the findings of psychological assessments with the data collected by medical imaging technologies. The end-to-end fusion architecture takes two separate types of data and integrates them. This makes the features more clear and the classification results more accurate. The model can pick up on little cross-modal interactions that standard unimodal or shallow fusion methods can't because it contains a trainable fusion layer that blends dense encoders for psychological assessments with convolutional neural networks for brain imaging.

REFERENCES

[1] Saravanan, V., Rajamani, A., Subramani, M., & Ramasamy, S. (2020). Exploring two-dimensional graphene and boron-nitride as potential nanocarriers for cytarabine and clofarabine anti-cancer drugs. *Computational Biology and Chemistry*, *88*, 107334.

[2] Subramanian, B., Saravanan, V., Nayak, R. K., Gunasekaran, T., & Hariprasath, S. (2019). Diabetic retinopathy-feature extraction and classification using adaptive super pixel algorithm. *International Journal on Engineering Advanced Technology*, *9*, 618-627.

[3] Kakani, T. A., Vedula, J., Mohammed, M., Gupta, R., Hudani, K., & Yuvaraj, N. (2025, June). Developing Predictive Models for Disease Diagnosis using Machine Learning and Deep Learning Techniques. In *2025 6th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)* (pp. 158-163). IEEE.

[4] Nithya, C., & Saravanan, V. (2018). A study of machine learning techniques in data mining. *Int. Sci. Refereed Res. J*, *1*, 31-38.

[5] Patil, S. C., Madasu, S., Rolla, K. J., Gupta, K., & Yuvaraj, N. (2024, June). Examining the Potential of Machine Learning in Reducing Prescription Drug Costs. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.

[6] Gupta, R., Kakani, T. A., Vedula, J., Mohammed, M., Hudani, K., & Yuvaraj, N. (2025, June). Advancing Clinical Decision-Making using Artificial Intelligence and Machine Learning for Accurate Disease Diagnosis. In *2025 6th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)* (pp. 164-169). IEEE.

[7] Rokicki, J., Wolfers, T., Nordhøy, W., Tesli, N., Quintana, D. S., Alnæs, D., ... & Westlye, L. T. (2021). Multimodal imaging improves brain age prediction and reveals distinct abnormalities in patients with psychiatric and neurological disorders. *Human brain mapping*, *42*(6), 1714-1726.

[8] Tian, Y. E., Di Biase, M. A., Mosley, P. E., Lupton, M. K., Xia, Y., Fripp, J., ... & Zalesky, A. (2023). Evaluation of brain-body health in individuals with common neuropsychiatric disorders. *JAMA psychiatry*, *80*(6), 567-576.

[9] Ghasemi, F., & Lino Modes Santos Machado, B. (2025). Medical Imaging Applications in the Diagnosis of Psychiatric Disorders. *Journal of Reviews in Medical Sciences*, *5*(1), 1-7.

[10] Shusharina, N., Yukhnenko, D., Botman, S., Sapunov, V., Savinov, V., Kamyshov, G., ... & Voznyuk, I. (2023). Modern methods of diagnostics and treatment of neurodegenerative diseases and depression. *Diagnostics*, *13*(3), 573.

[11] Mohammed, B. A., Senan, E. M., Rassem, T. H., Makbol, N. M., Alanazi, A. A., Al-Mekhlafi, Z. G., ... & Ghaleb, F. A. (2021). Multi-method analysis of medical records and MRI images for early diagnosis of dementia and Alzheimer's disease based on deep learning and hybrid methods. *Electronics*, *10*(22), 2860.

[12] Iyortsuun, N. K., Kim, S. H., Jhon, M., Yang, H. J., & Pant, S. (2023, January). A review of machine learning and deep learning approaches on mental health diagnosis. In *Healthcare* (Vol. 11, No. 3, p. 285). MDPI.