International Journal of Computer Vision and Signal Processing, 15(1), 41-48(2025) ORIGINAL ARTICLE

MNMD: A multimodal non-invasive mental disorder detection method

Rashed Mustafa*

Department of Computer Science and Engineering
University of Chittagong, Bangladesh

Mahir Shadid

Department of Computer Science and Engineering
International Islamic University Chittagong, Bangladesh

A. H. M. Sajedul Hoque

Department of Numerical Analysis
ELTE Eötvös Loránd University, Budapest, Hungary

Abstract

Millions of people worldwide suffer from mental health conditions like anxiety, stress, and depression, but early and precise detection is still difficult. This study introduces MNMD (Multimodal Non-Invasive Mental Disorder Detection), a system that uses an effective late fusion technique to combine results from the DASS-21 questionnaire with facial expressions in a unique way. In contrast to earlier methods, MNMD focuses on a low-complexity, non-invasive design that combines projected outputs using a "common-string" technique, lowering computational overhead while improving data resilience and variation. The system uses a variety of machine learning models and deep learning frameworks in addition to extensive image feature extraction using Gabor filters and facial landmark detection. With an impressive 98.43% accuracy rate, MNMD offers a quicker, privacy-preserving method of early mental health diagnosis while also demonstrating enhanced prediction performance and practicality for real-world implementation.

Contribution of the Paper: The primary contribution is the development of a novel, non-invasive multimodal fusion method called MNMD, which combines textual and visual data using a low-complexity late fusion technique and achieves higher accuracy (98.43%) in the detection of stress, anxiety, and depression.

Keywords: Common Mental Disorders, Multimodal Late Fusion, Fully-Connected Neural Network, Facial Landmark Detection, Feed Forward Neural Network, Bootstrap Sampling

IJCVSP

ISSN: 2186-1390 (Online)
http://cennser.org/IJCVSP

Article History:
Received: 15/3/2025
Revised: 10/72025
Accepted: 1/11/2025
Published Online: 23/11/2025

1. INTRODUCTION

Mental health disorders like anxiety, depression, and stress impact millions globally, affecting both emotional and physical well-being. Influenced by genetic, environmental, and psychological factors, these conditions cause symptoms such as persistent worry, sleep disturbances, and irritability. Early and accurate diagnosis is crucial for effective treatment, and machine learning offers a promising solution by integrating facial expressions and textual responses for more reliable detection. However, invasive algorithms raise privacy concerns, demand high computational resources, and may produce unreliable diagnoses. A simpler, more efficient approach is needed to enhance accuracy, user experience, and scalability in mental health care.

The main goal of this study is to provide a reliable,

*Corresponding author
Email addresses: rashed.m@cu.ac.bd (Rashed Mustafa), mahir.shadid@gmail.com (Mahir Shadid), sajed@inf.elte.hu (A. H. M. Sajedul Hoque)

multimodal late fusion method that uses a non-invasive procedure to precisely identify mental health issues like stress, anxiety, and depression. The goal is to improve this strategy, making it more efficient, successful, and suited to the complications of mental health condition recognition, by expanding on the work of [1]. The key objectives are as follows:

1. Patient Comfort: Ensures a seamless, non-invasive process for accessible and user-friendly mental health detection.
2. Facial Landmark Features: Enhances emotion detection by integrating landmark and Gabor features, improving accuracy under various conditions.
3. Efficient Data Fusion: Combines DASS-21 (Depression, Anxiety and Stress Scale-21) responses and facial features with minimal complexity while maintaining high accuracy.
4. Practical Validation: Undergoes extensive real-world testing to ensure reliability and accuracy in diverse scenarios.

2. RELATED WORKS

The application of machine learning in mental health detection has gained traction, particularly in leveraging unimodal methods such as textual data, EEG signals, or facial expressions. Priya et al. (2020) [2] utilized the DASS-21 survey to evaluate the severity of stress, depression, and anxiety, achieving an average accuracy of nearly 80% across classifiers like Decision Tree, SVM, Naive Bayes, KNN, and Random Forest. However, there are no mentions of hyperparameters that increases detection capacity of the models. Alshorman et al. (2022) [3] employed EEG spectrum analysis of the frontal lobes using Fast Fourier Transform (FFT) for feature extraction, coupled with SVM and Naive Bayes classifiers, reaching 98.21% accuracy in subject-wise classification. Although the FFT-based approach shows promise, it remains limited to stress detection through an invasive process. Jawad et al. (2023) [4] introduced the PS-CS optimization algorithm, a combination of particle swarm optimization and cuckoo search, to train CNNs for depression detection, achieving a remarkable 99.5% accuracy, outperforming traditional models like KNN and Decision Trees (69%-97%). However, relying solely on online posts to detect depression may introduce noise and unreliability, as social media posts often do not accurately reflect true mental health status. Al-Nafjan et al. (2024) [5] explored the underutilized potential of GSR signals for detecting anxiety, achieving classification accuracies of 96.9% and 98.2% using SVM, KNN, and Random Forest. However, its detection capability is restricted solely to anxiety. Fernandez et al. (2024) [6] examined stress detection using EEG signals and achieved significant improvements by incorporating simple statistical features into models such as LightGBM, CNN, KNN, and SVM. Despite these advancements, unimodal methods face limitations, including overreliance on subjective surveys, susceptibility to social desirability bias, and the inability to capture the multifaceted nature of mental health disorders. Visual data alone, while useful, often fail to represent the complex emotional and cognitive aspects of these conditions [7, 8, 9].

To address the shortcomings of unimodal approaches, multimodal methods have emerged as a more comprehensive solution, combining diverse data sources such as text, images, audio, and sensor data. These methods integrate multiple modalities to capture the complex interplay of factors influencing mental health. Park et al. (2022) [10] proposed a multimodal attention-based system utilizing text and speech data, achieving accuracies of 66% and 74%, respectively, through BERT-based embeddings. Although the model is non-invasive, it detects only depression through text and speech, while speech-based signals are susceptible to environmental noise interference, affecting reliability. Xie et al. (2022) [11] introduced a CNN-LSTM model that achieved 83.78% classification accuracy using video-based data from SDS and SAS tests, though environmental factors posed challenges. Mo et al. (2022) [12] explored integrating text and image data for anxiety detection, highlighting the potential of multimodal systems to improve diagnostic accuracy. Marriwala et al. (2023) [13] developed a hybrid deep learning model combining textual and audio features with CNN and Bi-LSTM, achieving superior performance on DAIC-WoZ data. Shadid et al. (2023) [1] advanced a non-intrusive late multimodal fusion method, combining text and image data using six machine learning techniques. For image data from KDEF and CK+ datasets, convolutional neural networks with real Gabor filters achieved up to 97.62% accuracy. Incorporating facial landmark and Gabor features proved effective in addressing environmental challenges, enhancing the reliability of image-based systems. The requirement for fixed-length input in fusion is unnecessary, limiting variance. The study (2024) [14] explores stress detection using wearable biosensors and the WESAD dataset, identifying EDA as the most significant signal. XGBoost achieves 98.8% accuracy and 98.7% F1-score using ACC, EDA, ECG, TEMP, and RESP, enabling a lightweight model for wearable biofeedback devices. Zhu et al. (2025) [15] propose MTNet, a multimodal transformer combining eye tracking and EEG for depression detection, achieving 91.79% accuracy with intermediate fusion yielding the best performance, which is also an invasive method. Park et al. (2025) [16] use ML models on VR therapy data to predict anxiety symptoms in SAD patients, with CatBoost achieving AUROCs of 0.852 (social phobia) and 0.866 (cognitive symptoms), and multimodal models outperforming unimodal ones. However, despite their potential, multimodal methods must account for user preferences, as invasive detection techniques can lead to feelings of discomfort, stigma, and embarrassment, deterring individuals from seeking help [17, 18, 19]. Additionally, over-engineering multimodal systems can intro-

duce unnecessary complexity, making them less practical for real-world applications.

3. PROPOSED METHOD

The process (Figure 1) begins with preprocessing text and image data, followed by splitting it into training, validation, and testing sets. To enable fusion, a lightweight "common-string" merging technique ensures dimensional alignment by assigning each prediction a unique key. Only outputs with matching keys are merged, preserving semantic consistency without complex feature alignment. After preprocessing, models including Convolutional Neural Networks (CNNs), Fully Connected Neural Networks (FCNNs), and eight machine learning algorithms: Naive Bayes, XGBoost, Decision Tree, Random Forest, LightGBM, CatBoost, Support Vector Machine (SVM), and k-Nearest Neighbors (K-NN) are applied. CNN and FCNN outputs, along with GridSearchCV-optimized classifiers, feed into a late fusion model using neural networks. The fused data is stored in a unified dataframe and processed through the model. Random Under-sampling (RUS) addresses class imbalance, and dropout layers in CNN, FCNN, and Feed Forward Neural Network (FFNN) reduce overfitting, enhancing generalization.

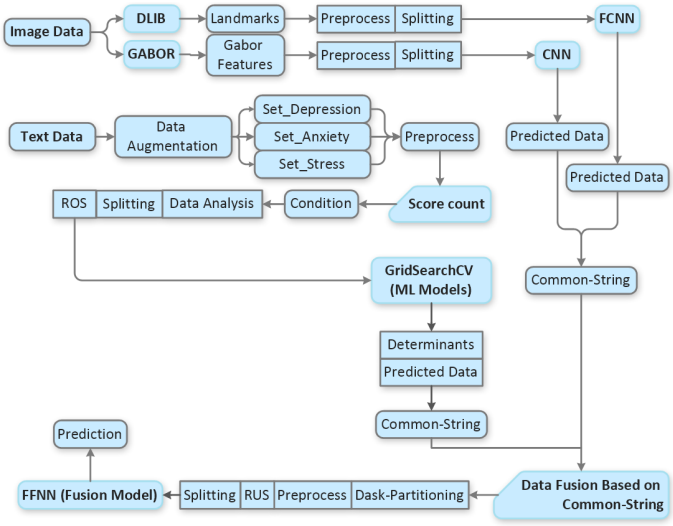


Figure 1: MNMD's workflow

3.1. *Image Data Processing Stage*

To further validate the method, individual models are evaluated based on their respective modalities in the results and discussion section. As noted earlier, the Convolutional Neural Network (CNN) handles image classification using Gabor features, while the Fully Connected Neural Network (FCNN) is used for classifying images based on facial landmark features.

3.1.1. *FCNN with Landmark-Based Features*

The Fully Connected Neural Network (FCNN) predicts mental disorders using facial landmark coordinates (x, y) extracted via the Dlib library. It comprises three dense layers with 128, 64, and 32 units, each followed by LeakyReLU activation (alpha = 0.1) to prevent neuron inactivity. Batch normalization stabilizes training, while dropout layers (0.4, 0.3, 0.2) and L2 regularization (0.01) mitigate overfitting. A softmax-activated output layer provides class probabilities. Training is optimized using EarlyStopping (after 10 stagnant epochs) and ReduceLROnPlateau (factor of 0.1) to improve convergence when validation accuracy plateaus.

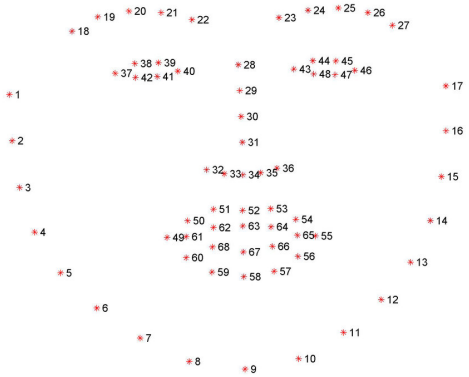


Figure 2: Facial landmark features

3.1.2. *CNN with Gabor-Based Filters*

The Convolutional Neural Network (CNN) model incorporates a Real-Gabor filter to extract texture, orientation, and frequency features while preserving spatial structure, ideal for facial expression detection (Figure 3). It includes three convolutional layers (6, 16, and 64 filters) with MaxPooling and Dropout (0.3) to reduce spatial dimensions and prevent overfitting. ReLU activation adds nonlinearity, and the output is flattened and passed through a fully connected layer with 128 units and another Dropout layer. The final softmax layer performs binary classification. EarlyStopping stops training after 10 stagnant epochs, while ReduceLROnPlateau lowers the learning rate by 0.1 when validation accuracy plateaus to improve convergence.

Figure 3: Before and After Gabor filtering

3.2. *Text Data Processing Stage*

Classifiers were chosen for complementary strengths, Decision Tree and Random Forest (interpretability, robust-

ness), k-Nearest Neighbors (local similarity), Support Vector Machine (high-dimensional modeling), XGBoost, LightGBM, CatBoost (efficient nonlinear learning), and Naive Bayes (sparse DASS-21 data). Hyperparameters (SVM's C and kernel, tree depths/min_samples, boosting learning rates/estimators, k-NN's n_neighbors) were optimized via GridSearchCV, and all DASS-21 scores were computed and features standardized to ensure uniform scaling.

3.3. ***Fusion Data Processing Stage***

In the fusion preparation stage (Algorithm 1), predicted image, text, and DASS-21 outputs are first aligned by appending "common strings," irrelevant features are dropped, and only matching records are merged. The resulting dataset is partitioned with Dask, balanced via random under-sampling, split into training, validation, and test sets, and then used to train the Feed Forward Neural Network (FFNN).

Algorithm 1: Fusion Data Processing Algorithm

Input: Predicted Gabor Data, Landmark Data, Text Data, DASS-21 Determinants
Output: Trained Feedforward Neural Network (FFNN) model

Step 1: Augment Data with Common Strings;
foreach *data point* **do**
  Append common strings to text and image data based on predicted labels;
end
Step 2: Feature Analysis and Selection;
Perform feature analysis and eliminate irrelevant or redundant features;
Step 3: Data Fusion using Common-string ;
Fuse datasets by performing a Cartesian product between data points where their common strings match;
Step 4: Preprocessing and Deduplication;
Apply preprocessing techniques and remove duplicates;
Step 5: Data Partitioning;
Partition the preprocessed data into manageable chunks using *Dask*;
Step 6: Feature-Target Separation;
Separate the data into independent features and target variables;
Step 7: Class Balancing;
Apply random under-sampling to balance class distributions;
Step 8: Data Splitting;
Split the balanced dataset into training, validation, and testing sets;
Step 9: Model Training;
Train the FFNN model using the fused and prepared data;

The fusion data preparation involves adding a "common string" key to each prediction record, aligning Gabor, landmark, and text-based outputs. This method ensures that only semantically relevant samples are matched, avoiding data loss and expensive feature alignment.

4. Results and Discussion

4.1. ***Dataset***

This study utilizes the CK+48 [20], KDEF [https://kdef.se/], and DASS-21 datasets [available on clinical platforms like https://www.healthfocuspsychology.com.au/tools/dass-21/]. The CK+48 dataset includes 981 image sequences for facial expression recognition, with participants ensuring clean skin and proper lighting. It covers seven expressions: anger, contempt, neutral, disgust, fear, happiness, and sadness. The KDEF dataset contains 2900 images with a similar collection process, except that images remain in RGB mode instead of being converted to grayscale.

(a) (b)

Figure 4: Samples of Facial Expression from CK+; a = Happy, b = Sad

(a) (b)

Figure 5: Samples of Facial Expression from KDEF; a = Happy, b = Sad

The DASS-21 dataset, comprising 581 data points, was collected by the author of [1, 2]. To augment the dataset, data augmentation technique was applied, increasing the sample size to 2000. This synthetic augmentation did not harm model performance; in fact, larger datasets led to improved results. The 21-indicators or determinants, the responses and the range of scores of the disorders are mentioned in the Table 1 and Table 2.

4.2. ***Data Preprocessing***

Text data analysis starts with preparing the data by handling null values and duplicates, followed by scoring based on participant responses collected via Google Forms in [1, 2]. Disorder classification (depression, anxiety, stress) is determined by the DASS-21 equation: ***score = sum of each determinant's rating points * 2***, with ratings

No.	Features	No.	Features
1	Found hard to wind down	11	Felt down-hearted and blue
2	Dryness of mouth	12	Getting agitated
3	Could not experience the positive feeling	13	Close to panic
4	Difficulty in breathing	14	Difficult to relax
5	Difficult to work up the initiative to do things	15	Unable to become enthusiastic
6	Overreact to situations	16	Aware of the action of the heart in the absence of physical exertion
7	Experience trembling	17	Felt was not worth much as a person
8	A lot of nervous energy	18	Intolerant to getting what I was doing
9	Nothing to look forward	19	Felt scared without any good reason
10	Worried about panic and making a fool of themselves	20	Felt life was meaningless
		21	Touchy

Table 1: Features of DASS-21 Dataset

Category	Questions
D (Depression)	Q3, Q5, Q10, Q13, Q16, Q17, Q21
S (Stress)	Q1, Q6, Q8, Q11, Q12, Q14, Q18
A (Anxiety)	Q2, Q4, Q7, Q9, Q15, Q19, Q20

Table 2: DASS-21 Questions Distribution

ranging from 0 (not applicable) to 3 (very much applicable).

For image data, KDEF images are resized to 48x48 pixels, converted to grayscale, and normalized using the Gabor filter. Dlib's face detector is used for landmark detection. Emotion remapping links emotions to their corresponding disorders. The preprocessed data is stored in a dataframe and divided into three subsets: training, testing, and validation. The model is trained on the training set and evaluated using the testing and validation sets.

*4.3. **Experiment Results***

The DASS-21 dataset was augmented via bootstrap sampling to balance classes and increase variance with custom-noise synthetic samples, validated its suitability, and optimized model hyperparameters using GridSearchCV. For image data, Gabor filters extract sharp edges, while dlib detects 68 facial landmarks to capture detailed regional features.

Condition	Depression	Anxiety	Stress
Normal	0-9	0-7	0-14
Disorder	≥ 10	≥ 8	≥ 15

Table 3: Scoring and Labeling for Depression, Anxiety, and Stress

4.3.1. Outcomes of the suggested MNMD model (FFNN) on fusion data

Before training the FFNN model, the text and image data were combined using the proposed MNMD method. The FFNN model was then trained on this fused data, functioning as the MNMD model. The table (Table 4) shows the classification performance of the FFNN-MNMD model, achieving high F1 scores (96- 99%) in the depression, anxiety, stress and normal states. With an overall accuracy of 98.43%, the model demonstrates excellent reliability in distinguishing these mental health states.

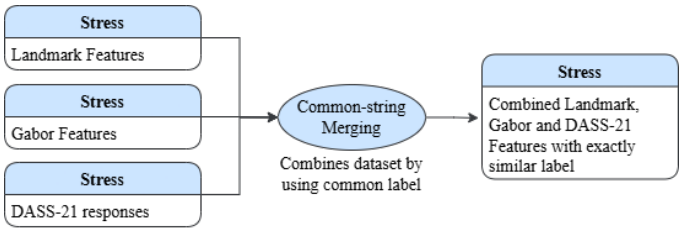


Figure 6: Feature fusion technique of MNMD: The merging technique aligns the data points together only if the target label (common-string) matches

Class	Precision	Recall	F1-Score
Depression	97%	96%	97%
Anxiety	98%	99%	98%
Stress	98%	95%	96%
Normal	98%	100%	99%
Accuracy		98.43%	

Table 4: Classification Report of FFNN-MNMD

K=10 folds cross-validation is commonly used as it balances bias and variance, offering reliable performance estimates (Table 5) with manageable computational cost.

4.3.2. Outcomes of GridSearchCV with ML Models, CNN-FCNN on images with Real-Gabor Filter and DLib

In this research, among the eight machine learning models tested (Naive Bayes, XGBoost, Decision Tree, Random Forest, SVM, LightGBM, CatBoost, and K-NN), GridSearchCV identified Support Vector Machine (SVM) as the best-performing model (Table 6).

The images from the CK+ dataset originally have a resolution of 48x48 pixels, while KDEF images are colored with a resolution of 562x762 pixels. To combine these

Fold	Validation Accuracy	Validation Loss
1	0.9651	0.2990
2	0.9822	0.2124
3	0.9722	0.2375
4	0.9942	0.1697
5	0.9869	0.1972
6	0.9844	0.2439
7	0.9715	0.2757
8	0.9850	0.2105
9	0.9422	0.3337
10	0.9801	0.2149
Average	**0.9851**	**0.2270**

Table 5: K-Fold Cross Validation Results for the MNMD Model

Metric	Before Tuning	After Tuning
Accuracy	90.50%	99.25%
Error Rate	9.50%	0.75%
Precision	90.55%	99.26%
Recall	90.50%	99.25%
F1 Score	90.44%	99.25%

Table 6: SVM (Best Model) Performance Before and After Hyperparameter Tuning using GridSearchCV

datasets, KDEF images were converted to grayscale and resized to 48x48 pixels. The combined dataset was then used to train CNN and FCNN models. The research categorized emotions into Positive and Negative arousal classes (Table 7).

4.3.3. ***Suggested Approach Vs Existing Approaches***

This subsection presents two key findings from the research. The first, and most significant, is a comparison of the proposed approach with existing methods. The second is a comparison between the machine learning algorithms utilized, which were fine-tuned with hyperparameters, and the methodologies employed in previous studies [2, 1].

MNMD vs Existing Approaches:

The Multimodal Data Late Fusion model demonstrated excellent performance compared to other existing algorithms. Its simplicity and straightforward design contributed to achieving an impressive accuracy of **98.43%**.

The study used a Z-test to compare (Table 8) the accuracies of two models, determining whether the observed difference in correct predictions is statistically significant. This test is suitable when the models have different class distributions and accuracy is the only available metric. It helps researchers assess whether the accuracy difference is due to random chance or reflects a true difference in model performance. The comparison table (Table 9) shows the broader spectrum covered by MNMD approach.

Model	Class	Precision	Recall	F1-Score	Acc.
CNN	Negative	91%	96%	94%	92%
	Positive	93%	84%	88%	
FCNN	Negative	73%	93%	82%	80%
	Positive	91%	67%	77%	

Table 7: Classification Report for CNN-Gabor Filter and FCNN-DLib Models

Metrics	TI-Fusion	MNMD
Accuracy	97.62%	98.43%
Number of samples	4481	5504
Z-statistic	2.906	
P-value	0.0036	
Statistically significant difference in accuracy ($p < 0.05$)		

Table 8: Comparison using Z-test (The probability value ($p < 0.05$) is the standard threshold and it is also used in comparison for TI-Fusion [1])

Feature	TI-Fusion	MNMD
Accuracy	97.62%	98.43%
Disorder Coverage	Anxiety	Depression, Anxiety, Stress
DASS-21 Feature	7	21
Image Features	Gabor Features	Gabor and Landmark Features

Table 9: Comparison table of the core features and performance metrics of the TI-Fusion [1] and MNMD model in mental disorder detection, highlighting improvements in accuracy, disorder coverage, and feature utilization

The bootstrapping method with a custom distribution increased the DASS-21 sample size to 5501. Landmark detection skipped samples where no faces were detected. The two-proportion Z-test results show a significant accuracy difference between the MNMD model (98.43%) and the TI-Fusion [1] model (97.62%), with a Z-statistic of 2.906 and a P-value of 0.0037 ($p < 0.05$). This suggests that MNMD is more effective for accurate predictions, supported by the larger sample size of 5504.

Comparing based on claims: The research [1] has the claim that:

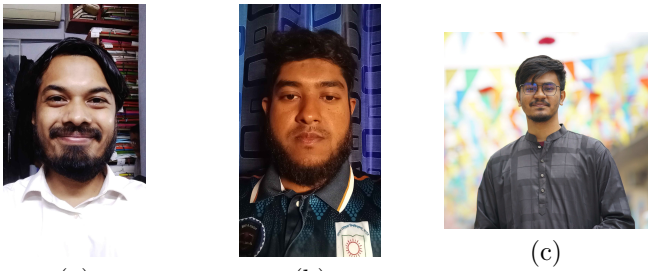
- *Equalized test/predicted data size is necessary:* TI-

Fusion requires equal-sized datasets, but common-key-based fusion does not. Keeping datasets untrimmed enhances variety, akin to a *Cartesian Product* approach. The Cartesian fusion method generates all possible paired combinations across modalities where the common strings match. This increases data variance by exposing the model to diverse cross-modal interactions. Overfitting is mitigated through regularization techniques such as dropout layers and L2 penalties in the FFNN, and balancing techniques like random under-sampling. Together, these steps ensure that increased variance improves generalization without compromising model stability.

- *Naive Bayes has no parameters:* Contrary to the claim that Naïve Bayes has no parameters, the "var_ smoothing" hyperparameter in GaussianNB stabilizes variance, improving accuracy.

4.4. ***Real-world Testing***

To test the fusion model, real-world testing was conducted by collecting image (Figure 7) and text data (Table 10) from few individuals experiencing depression, anxiety, and stress across professions, with their consent, via Google Forms. The result of the test is shown in the Table 11. This approach relies on practical data analysis to validate models and support data-driven decision-making.



(a) (b) (c)

Figure 7: Sample images of subjects used in the study. (a) Subject-1, (b) Subject-2, (c) Subject-3.

4.5. ***Enhancements***

This section presents enhancements to the MNMD, building on conventional approaches to address their limitations. These improvements aim to further optimize the model's accuracy, efficiency, and overall effectiveness in classification tasks.

1. Non-Invasive: Unlike invasive methods, MNMD ensures non-intrusive detection.
2. Improved Data Variance: Integrates landmark and Gabor features while preserving predicted data for training. Data augmentation enhances the DASS-21 dataset.
3. Broader Prediction: Accurately classifies four severity levels; depression, anxiety, stress, and normality without needing advanced systems.

Question	Subject-1	Subject-2	Subject-3
Q1 (Stress)	0	1	0
Q2 (Anxiety)	0	1	1
Q3 (Depression)	0	1	1
Q4 (Anxiety)	0	1	0
Q5 (Depression)	1	1	1
Q6 (Stress)	0	0	2
Q7 (Anxiety)	0	1	0
Q8 (Stress)	0	1	0
Q9 (Anxiety)	0	1	0
Q10 (Depression)	0	1	0
Q11 (Stress)	0	0	2
Q12 (Stress)	0	1	1
Q13 (Depression)	0	1	0
Q14 (Stress)	0	1	1
Q15 (Anxiety)	0	1	0
Q16 (Depression)	0	1	1
Q17 (Depression)	0	1	1
Q18 (Stress)	0	1	0
Q19 (Anxiety)	0	0	0
Q20 (Anxiety)	0	1	0
Q21 (Depression)	0	1	0

Table 10: DASS-21 responses from three subjects across stress, anxiety, and depression indicators.

Category	Subject-1	Subject-2	Subject-3
Depression	3.43%	98.91%	40.48%
Anxiety	3.90%	0.09%	4.18%
Stress	1.10%	0.46%	13.68%
Normality	91.56%	0.53%	41.64%
Disorder	None	Depression	None

Table 11: MNMD's Prediction of Depression, Anxiety, Stress, and Normality percentage for the subjects

4. Low Complexity: Simplifies design by avoiding data trimming, with FCNN used only for landmark features.
5. Efficient Computation: Uses *dask partitioning* to optimize training on standard hardware (Ryzen 5 3500U, 10GB RAM). Despite handling 173,904 data points, processing remains under a minute, enabling use on lower-end devices.

5. CONCLUSIONS

In this study, MNMD is introduced, a multimodal non-invasive approach for identifying mental health issues that combines data from the DASS-21 questionnaire with facial expressions. The system's exceptional detection rate of 98.43% was attained by employing a late fusion method that combined textual indicators with landmark-based and

Gabor-based image features, exceeding a number of existing techniques. A thorough validation process using k-fold cross-validation and practical evaluation validated the model's efficacy and dependability across stress, anxiety, and depression classes. MNMD is also useful for real-world mental health diagnosis since it tackles issues like class imbalance, computational efficiency, and data complexity. Even though MNMD takes a non-invasive approach, there may be biases introduced by using facial expression data. Changes in illumination, camera quality, and participant demographics (e.g., age, skin tone, or facial structure) can impact the precision of Gabor feature extraction and landmark detection. To guarantee equal performance, future research should employ techniques like adaptive preprocessing and bias prevention, as well as fairness assessments across a range of demographics. To avoid abuse, concerns about privacy must also be resolved by safe data handling, informed consent, and on-device inference.

References

[1] M. Shadid, M. S. Afnan, M. J. A. Patwary, Ti-fusion: A multimodal anxiety disorder detection method, in: 2023 6th International Conference on Electrical Information and Communication Technology (EICT), 2023, pp. 1–6. doi:10.1109/EICT61409.2023.10427924.

[2] A. Priya, S. Garg, N. P. Tigga, Predicting anxiety, depression and stress in modern life using machine learning algorithms, Procedia Computer Science 167 (2020) 1258–1267.

[3] O. AlShorman, M. Masadeh, M. B. B. Heyat, F. Akhtar, H. Almahasneh, G. M. Ashraf, A. Alexiou, Frontal lobe real-time eeg analysis using machine learning techniques for mental stress detection, Journal of integrative neuroscience 21 (1) (2022) 20.

[4] K. Jawad, R. Mahto, A. Das, S. U. Ahmed, R. M. Aziz, P. Kumar, Novel cuckoo search-based metaheuristic approach for deep learning prediction of depression, Applied Sciences 13 (9) (2023) 5322.

[5] A. Al-Nafjan, M. Aldayel, Anxiety detection system based on galvanic skin response signals, Applied Sciences 14 (23) (2024) 10788.

[6] J. Fernandez, R. Martinez, B. Innocenti, B. Lopez, Contribution of eeg signals for students' stress detection, IEEE Transactions on Affective Computing (01) (2024) 1–12.

[7] K. R. Scherer, Emotions are emergent processes: they require a dynamic computational architecture, Philosophical Transactions of the Royal Society B: Biological Sciences 364 (1535) (2009) 3459–3474.

[8] G. Lensvelt-Mulders, Surveying sensitive topics, in: International handbook of survey methodology, Routledge, 2012, pp. 461–478.

[9] M.-T. Wang, D. A. Henry, C. L. Scanlon, J. Del Toro, S. E. Voltin, Adolescent psychosocial adjustment during covid-19: an intensive longitudinal study, Journal of Clinical Child & Adolescent Psychology (2022) 1–16.

[10] J. Park, N. Moon, Design and implementation of attention depression detection model based on multimodal analysis, Sustainability 14 (6) (2022) 3569.

[11] W. Xie, C. Wang, Z. Lin, X. Luo, W. Chen, M. Xu, L. Liang, X. Liu, Y. Wang, H. Luo, et al., Multimodal fusion diagnosis of depression and anxiety based on cnn-lstm model, Computerized Medical Imaging and Graphics 102 (2022) 102128.

[12] H. Mo, Y. Li, S. Yang, W. Zhang, S. Ding, Sff-da: Sptialtemporal feature fusion for detecting anxiety nonintrusively, arXiv preprint arXiv:2208.06411.

[13] N. Marriwala, D. Chaudhary, et al., A hybrid model for depression detection using deep learning, Measurement: Sensors 25 (2023) 100587.

[14] G. Mazumdar, K. Singh, D. Sethia, Ml-based stress classification and detection using multimodal affect dataset, in: Advances in Networks, Intelligence and Computing, CRC Press, pp. 210–221.

[15] F. Zhu, J. Zhang, R. Dang, B. Hu, Q. Wang, Mtnet: Multimodal transformer network for mild depression detection through fusion of eeg and eye tracking, Biomedical Signal Processing and Control 100 (2025) 106996.

[16] J.-H. Park, Y.-B. Shin, D. Jung, J.-W. Hur, S. P. Pack, H.-J. Lee, H. Lee, C.-H. Cho, Machine learning prediction of anxiety symptoms in social anxiety disorder: utilizing multimodal data from virtual reality sessions, Frontiers in Psychiatry 15 (2025) 1504190.

[17] P. W. Corrigan, A. Kerr, L. Knudsen, The stigma of mental illness: Explanatory models and methods for change, Applied and Preventive Psychology 11 (3) (2005) 179–190. doi:https://doi.org/10.1016/j.appsy.2005.07.001.
URL https://www.sciencedirect.com/science/article/pii/S0962184905000053

[18] C. Vidal, M. Garcia, L. Benito, N. Milà, G. Binefa, V. Moreno, Use of text-message reminders to improve participation in a population-based breast cancer screening program, Journal of medical systems 38 (9) (2014) 118.

[19] G. Rubeis, F. Steger, A burden from birth? non-invasive prenatal testing and the stigmatization of people with disabilities, Bioethics 33 (1) (2019) 91–97.

[20] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression, in: 2010 ieee computer society conference on computer vision and pattern recognition-workshops, IEEE, 2010, pp. 94–101.