



Multimodal autism detection: Deep hybrid model with improved feature level fusion



S. Vidivelli*, P. Padmakumari, P. Shanthi

Department of Computer Science and Engineering, School of Computing, SASTRA Deemed to be University, Thanjavur, Tamilnadu, 613402, India

ARTICLE INFO

Keywords:

Multimodal autism detection
Improved singular spectrum entropy
Proposed shape local binary texture
Improved active appearance model
Hybrid classifier

ABSTRACT

Objective: Social communication difficulties are a characteristic of autism spectrum disorder (ASD), a neuro-developmental condition. The earlier method of diagnosing autism largely relied on error-prone behavioral observation of symptoms. More intelligence approaches are in progress to diagnose the disorder, which still demands improvement in prediction accuracy. Furthermore, computer-aided design systems based on machine learning algorithms are extremely time-consuming and difficult to design. This study used deep learning techniques to develop a novel autism detection model in order to overcome these problems.

Methods: Preprocessing, Features extraction, Improved Feature level Fusion, and Detection are the phases of the suggested autism detection methodology. First, both input modalities will be preprocessed so they are ready for the next stages to be processed. In this case, the facial picture is preprocessed utilizing the Gabor filtering technique, while the input EEG data is preprocessed through Wiener filtering. Subsequently, features are extracted from the modalities, from the EEG signal data, features like Common Spatial Pattern (CSP), Improved Singular Spectrum Entropy, and correlation dimension, are extracted. From the face image, features like the Improved Active Appearance model, Gray-Level Co-occurrence matrix (GLCM) features and Proposed Shape Local Binary Texture (SLBT), as well are retrieved. Following extraction, enhanced feature-level fusion is performed to fuse the features. Ultimately, the combined features are fed into the hybrid model to complete the diagnosis. Models such as Convolutional Neural Networks (CNN) and Bidirectional Gated Recurrent Units (Bi-GRU) are part of the hybrid model.

Results: The suggested MADDHM model achieved an accuracy of about 91.03 % regarding EEG and 91.67 % regarding face analysis meanwhile, SVM=87.49 %, DNN=88.59 %, Bi-GRU=90.02 %, LSTM=87.49 % and CNN=82.02 %.

Conclusion: As a result, the suggested methodology provides encouraging outcomes and opens up possibilities for early autism detection. The development of such models is not only a technical achievement but also a step forward in providing timely interventions for individuals with ASD.

1. Introduction

A computational intelligence approach called "bio-inspired computation" uses biological system models or principles to solve challenging real-world issues like autism detection. So this journal is well suited for this research model for effective analysis and generates the best solution for the problems associated with the autism diagnosis. Autism is classified as a neuro-developmental disorder that significantly impacts a person's social development in both childhood and adulthood. It has typically happened as a result of environmental variables or any hereditary connections, which affect children's and adults' social and

cognitive abilities in addition to their neural systems. Its symptoms vary widely in terms of both their severity and scope [1]. Frequent symptoms of the disorder include repetitive behaviors, obsession interests, and difficulty communicating, especially in social contexts [2]. Several studies have been conducted to detect ASD in children [3]. In addition to being crucial for computers to comprehend human states in BCI, emotion recognition has numerous potential uses in a variety of other domains [4]. Since early identification and treatment of ASD can somewhat reduce symptoms, it improves the person's overall quality of life. Nevertheless, a great deal of valuable time might be spent while diagnosing ASD because it cannot be accurately identified by displaying

* Corresponding author.

E-mail address: vidivelli@cse.sastra.edu (S. Vidivelli).

just the behaviors of adults or children in a clinic [5]. Studies have been conducted on infants to adults for the early detection of ASD including modalities like facial expressions [6], emotion identification, and eye-tracking mechanisms.

Nonetheless, it has been noted that children with ASD have trouble processing faces. The accuracy and latency of emotion detection in children with ASD and normally developing children were examined while seeing recordings of faces going from a neutral expression to one of the six fundamental emotions [7]. This is due to the fact that multiple studies [8,9] discovered that, in comparison to healthy children of the same age, children with ASD had trouble identifying and comprehending facial expressions. Meanwhile, the prediction of ASD using signals [10,11] is gaining attraction due to its ability to distinguish between normal and abnormal children. With its exceptional temporal resolution and accessibility in most situations, EEG is a reasonably priced approach. The anatomical and functional networks of the human brain have improved because of the use of EEG data [12]. With this advent of machine learning [13] and Deep Learning (DL) algorithms [14], ASD Prediction has reached a new high in early diagnosis. In machine learning, neural networks used for deductive reasoning, particularly in medical decision-making, can face significant challenges. The performance of such models is highly dependent on the quality of the input data, as poor data quality can adversely affect the accuracy and reliability of the network's decisions [15].

Moreover, the deep learning model, particularly, CNN makes use of the temporal nature of data to construct non-linear simulations of connections in input data [16,17]. Furthermore, CNN can effectively simulate the spatial dependence of the data [18,19]. Each computational layer of the network [20,21], as in the preceding layer, would be made up of several feature maps. It can only identify one emotion from the signal, and the emotion that is taken out of the signal is typically a combination of multiple emotions [22]. Additionally, explores the application of transfer learning schemes for the detection of autism, demonstrating that leveraging pre-trained models can enhance diagnostic accuracy and reduce the need for extensive training data. However, transfer learning can lead to overfitting if the pre-trained features are not well-suited to the nuances of the new task. [23], the early identification of symptoms associated with Autism Spectrum Disorder (ASD) holds promise for mitigating its impact. However, diagnosing ASD is challenging due to overlapping symptoms with other mental health conditions, leading to instances of misdiagnosis. Yet, if ML models could elucidate the rationale behind their ASD predictions, healthcare professionals could make more informed decisions during early assessment [24]. This served as the driving force behind our research, as early ASD detection can lead to timely intervention, thereby improving patient and family outcomes.

1.1. Motivation

Traditional methods for autism detection often involve extensive manual feature engineering, which is not only time-consuming but also prone to human bias. The quality of these methods heavily relies on the expertise of individuals selecting and designing features, which can impact the reliability of the results. In the context of structural MRI, existing approaches tend to focus more on distinguishing patterns between ASD and control groups rather than addressing the core classification problem. Furthermore, when dealing with small sample sizes, these methods often yield unreliable results due to poor generalization, which deteriorates accuracy as dataset sizes increase. The accuracy of these approaches is also affected by the variability in imaging protocols, leading to inefficient generalization. Additionally, many existing methods focus predominantly on functional findings while neglecting crucial structural information. To tackle these challenges, this paper suggests a new multimodal autism detection framework that integrates advanced preprocessing techniques, diverse feature extraction, and a hybrid model approach. The proposed framework aims to significantly

enhance the accuracy of autism detection. This approach holds the potential for more reliable and earlier diagnoses, thus improving the overall effectiveness of autism detection methodologies.

1.2. Contribution

The major contributions are as follows:

1. Proposing the Multimodal Autism Detection with Deep Hybrid Model (MADDHM), that considers the application of both facial and EEG modalities.
2. Employing the Improved Singular Spectrum Entropy feature, the Improved Active Appearance model-based feature, and the proposed SLBT feature along with the other feature to effectively capture complex patterns can lead to more accurate and reliable detection.
3. Deploying an improved feature-level fusion approach to fuse the feature set from both modalities ensures accurate diagnosis and significantly provides a more holistic view of the subject's condition.
4. Proposing a hybrid model that integrates Convolutional Neural Networks (CNN) and Bidirectional Gated Recurrent Unit (Bi-GRU) to enhance the model's robustness and detection performance.

1.3. Organisation summary

The paper is formatted as: The review is covered in Section 2. Section 3 provides an overview of the multimodal autism detection procedure. Section 4 determines the pre-processing, feature extraction, and Improved feature-level fusion of the developed model. Section 5 depicts the detection procedure using a hybrid classifier and its objective function. The findings and conclusion are presented in Section 6.

2. Literature review

In 2020, Eni et al. [25] has retrieved prosodic, acoustic, and conversational elements from voice recordings of Hebrew-speaking youngsters who took an Autism Diagnostic Observation Schedule (ADOS) evaluation. A number of Deep Neural Network (DNN) algorithms were developed, their performance was evaluated against both Linear Regression (LR) and Support Vector Regression (SVR) models, and these features were then used to predict ADOS results. They found that the best results came from a CNN. Using multiple subsamples of the available data, this technique correctly predicted ADOS scores, with a mean correlation of 0.72 and a mean Root Mean Square Error (RMSE) of 4.65. The FC-DNN model 3 had the most steady performance among the 50 dataset selections, with mean correlations of lower (mean $R = 0.70 \pm 0.09$) and maximum RMSE values (mean = 4.95 ± 0.65). The correlation standard deviation among the CNN and FC-DNN models was 0.09. The FC-DNN provides better ADOS SCORES with mean RMSE and a mean correlation. Nevertheless, the approach has limited availability of data and offers a high cost of computation.

In 2022, Barik et al. [25] has used to identify biomarkers of autism in young children. They concentrated on Power Spectral Density (PSD) and Preferred Phase Angle (PPA) of brain oscillations. When it came to PPA data, the ML-based classifier's classification accuracy was greater (88 %) than it was for PSD characteristics (82 %). Additionally, they used a novel fusion method that integrated PSD and PPA information to achieve an average classification accuracy. The classifiers obtain 94 % and 98 % classification accuracy, respectively, using spectral-domain analysis, making score-level fusion and somewhat more accurate classification than feature-level fusion. However, gathering magnetoencephalography (MEG) data from an identical child is a challenging effort that must be done in order to evaluate the individual achievements of both intrinsic and task-driven nature.

In 2021, Vakadkar et al [27] has created an ML technique for detecting ASD in children. Presently, clinical standardized tests are the only tools used to diagnose ASD. This results in not only a delay in

detection but also a significant rise in medical expenses. Machine learning methods were combined to conventional procedures to increase diagnosis time and accuracy. According to the findings, LR has the best accuracy for the chosen dataset. However, the lack of substantial open-source ASD datasets was the main research constraint.

In 2019, Ramírez-et al [28] has discovered that Child-Robot Interaction (CRI) and technologically-based methods for automated behavior assessment can increase the early detection of ASD indications in naturalistic behavioral observation. Additionally, by combining automated video coding and examining children's behavioral responses, computer vision could assist clinicians in diagnosing ASD more quickly. Additionally, the supervised behavior of ONO, a low-cost modular robot was successfully modulated by the feedback information regarding the child's performance, enhancing the CRI and the kids' visual attention. This study has shown the viability of detecting and measuring variations in the behavioral patterns triggered by the CRI between TD children and Children with Autism Spectrum Disorder (CwASD). The approach has the feasibility of quantifying and identifying differences in the patterns of TD children behavior however, the model failed when the child's hands created occlusion.

In 2022, Saranya et al [29] suggested a new diagnosis software that blends unique fuzzy hybrid DCNN and fusion of face expressions and human gaits utilizing input video sequences. Considering an optimal accuracy of 95 %, the developed deep learning model outperforms another state-of-the-art method with up to as 30 % on prediction accuracy. However, bio-inspired fuzzy optimizers are required to enhance the suggested approach in order to decrease dimensionality.

In 2022, Lu et al [30] proposed employing predictions based on combined resting state MRI and genetic data to diagnose ASD. This study employed fMRI and gene expression information from 71 participants in the National Database for Autism Research (NDAR). T-test and Support Vector Machine-Recursive Feature Elimination (SVM-RFE) were used for best feature selection and feature reduction. The effectiveness of the method is supported by the experimental results. Moreover, earlier studies have demonstrated a strong correlation between ASD and the genetic and imaging characteristics were retrieved. The suggested approach produced a classification accuracy of 83.6 %, surpassing that of any classifier using feature integration or single-modal features. The suggested model offers better predictive and accuracy results. However, the motion-artifact can further degrade the quality of physiological signals, leading to inaccurate measurements.

In 2020, Puli et al [31] has presented a method for detecting and mitigating the consequences of physical exercise in real-time. The proposal was for a novel Kalman-like filter to integrate accelerometry and heart rate data. The filter keeps track of the user's pulse rate under different mobility assumptions and chooses the most appropriate anxiety detection model based on the user's motion circumstances. A test of the method utilizing data from a sample of children with ASD reveals a considerable reduction in false positives and an overall accuracy of 93 % for detecting arousal, compared with the state-of-the-art. However, investigation of the system performance in naturalistic settings was challenging.

In 2022, Devika et al [32] has presented an outlier identification strategy for Structural Magnetic Resonance Imaging (sMRI)-based ASD diagnosis. A Generative Adversarial Network (GAN) is trained only on sMRI pictures of healthy patients in order to discover spatio-temporal patterns in structural brain connections. The GAN generator reconstructs the subsequent three adjacent slices given an input stack of three adjacent slices. It is demonstrated that the coronal modality more effectively encodes structural data for the diagnosis of ASD, and that the metrics used in model training and reconstruction loss computation affect the detection efficiency. There was a need to investigate the alternative architectures to SAGAN, such as Dense-Attentive GAN, for improving the detection performance.

In 2022, Alsaade, F.W. et al [33] has suggested that communities and psychiatrists identify autism using experimental means based on face

Table 1
Existing models based on ASD detection model.

Author	Method	Features	Drawbacks
Eni et al. [25]	FC-DNN model	<ul style="list-style-type: none"> Better ADOS SCORES with mean RMSE and a mean correlation 	<ul style="list-style-type: none"> Limited availability of data and high cost of computation.
Barik et al. [26]	ML-based classifier	<ul style="list-style-type: none"> Greater specificity, sensitivity, and classification accuracy 	<ul style="list-style-type: none"> Individual contributions of both intrinsic and task-driven nature must be investigated by collecting Magneto-encephalography (MEG) data from the same child
Vakadkar et al. [27]	ML techniques	<ul style="list-style-type: none"> Highest accuracy, Precision, recall, and F1-score 	<ul style="list-style-type: none"> The primary limitation of this study was the scarcity of large and open-source ASD datasets
Ramírez et al. [28]	ML neural models	<ul style="list-style-type: none"> Feasibility of quantifying and identifying differences in the patterns of TD children behavior 	<ul style="list-style-type: none"> The program failed when the child's hands created occlusion.
Saranya et al. [29]	fuzzy hybrid deep CNN	<ul style="list-style-type: none"> Better prediction time, accuracy, sensitivity, specificity, and less RMSE error 	<ul style="list-style-type: none"> Need for bio-inspired fuzzy optimizers to improve the proposed algorithm to reduce dimensionality
Lu et al. [30]	Multi-kernel SVM	<ul style="list-style-type: none"> Better predictive models, feasibility, and higher accuracy 	<ul style="list-style-type: none"> Motion-artifacts can further degrade the quality of physiological signals, leading to inaccurate measurements. Still required to investigate system performance in naturalistic settings.
Puli et al. [31]	modified Kalman filter	<ul style="list-style-type: none"> Higher sensitivity, specificity, and accuracy 	<ul style="list-style-type: none"> Alternative architectures to SAGAN, such as Dense-Attentive GAN, must be investigated
Devika et al. [32]	GAN	<ul style="list-style-type: none"> Higher accuracy and improves detection performance 	<ul style="list-style-type: none"> Computational complexity and requires more memory Susceptible to overfitting when dealing with limited data
Alsaade, F. W. et al. [33]	VGG19	<ul style="list-style-type: none"> Better prediction accuracy 	<ul style="list-style-type: none"> The efficiency of 1D-CNN heavily relies on the quality and quantity of the sequential data which might impact the accuracy of the model.
Akter, S. et al [34]	ResNet50	<ul style="list-style-type: none"> Higher precision rate and consumes lesser time for execution. 	
Kareem et al. [35]	1D-CNN	<ul style="list-style-type: none"> This method offers robust feature extraction and effectively identifies the crucial feature indicative of ASD 	
Awaji et al. [36]	Deep learning techniques	<ul style="list-style-type: none"> Combining features from multiple CNNs provides a more comprehensive analysis of facial characteristics. 	<ul style="list-style-type: none"> The complexity of integrating features from multiple CNNs increased the risk of overfitting, particularly if the model was trained on a limited dataset.

traits. It does this by utilizing a simple web application built on a deep learning system, namely a convolutional neural network with transfer learning and the Flask framework. Group Visual Geometry Exception. The pre-trained models used for the categorization challenge were NASNET Mobile and Network (VGG19). There were 2940 face photos made up of the dataset that was gathered from the Kaggle platform and utilized to test these models. All three deep learning models were

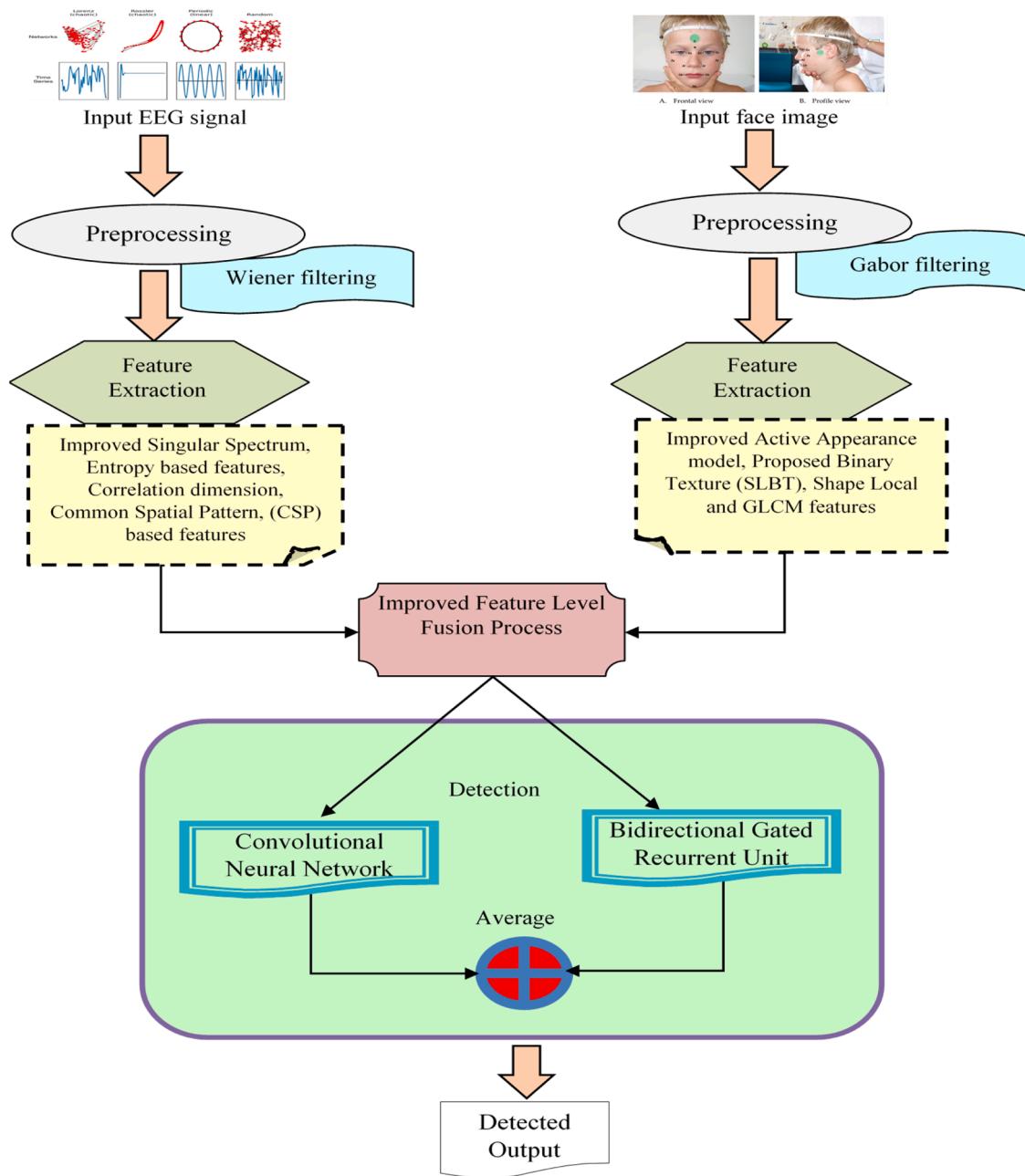


Fig. 1. Overview of proposed methodology.

evaluated using standard assessment measures, including sensitivity, specificity, and accuracy. The model with the greatest accuracy result, the Xception model, came in at 91 %, followed by VGG19 (80 %) and NASNET Mobile (78 %). However, the autism detection based on the VGG19 model was computationally complex and required more memory.

In 2023, Akter, S. et al [34] has compared CPU using state-of-the-art neural network algorithms, including Xception, Mobilenet, Resnet50, Inceptionv3, VGG16, XGBOOST-VGG16, and our proposed models and graphics processing unit resources. Using facial photos of autistic and non-autistic people, we developed a system for categorizing autism. In every test that was done, the GPU was found to perform better than the CPU. Furthermore, compared to the CPU, the GPU performance of the neural network models improved in terms of accuracy. The suggested model offers a higher precision rate and consumes less time to execute the detecting result. However, the suggested method was susceptible to

over-fitting when dealing with limited data.

In 2023, Kareem et al. [35] has proposed a 1D-CNN model specifically for analyzing sequential data related to autism. The study emphasized the model's capability to effectively capture temporal patterns and trends in such data. The 1D-CNN was used to identify and learn significant features indicative of ASD, addressing the need for robust feature extraction methods in autism detection. The research highlighted the 1D-CNN's ability to learn and generalize from the data, demonstrating its potential for enhancing diagnostic accuracy.

In 2023, Awaji et al. [36] has proposed a hybrid model that combines multiple CNN-based feature extraction techniques for analyzing facial images. This approach aimed to leverage the strengths of various CNN architectures to enhance the detection of subtle facial features associated with ASD. The results demonstrated that combining CNN features led to improved accuracy in detecting ASD, highlighting the potential of this method for early diagnosis. However, the complexity of integrating

features from multiple CNNs increased the risk of overfitting, particularly if the model was trained on a limited dataset.

The ASD detection model is shown in Table 1.

Because of the large variability among people, diagnosing ASD by clinical examination (cognitive tests) is difficult. The FC-DNN model [25] delivers improved ADOS SCORES with a mean RMSE and a mean correlation; nevertheless, these approaches require high-quality data recordings from children with a range of autism severity levels, but its availability is limited and charges high for the computation process. Individual contributions of both intrinsic and task-driven characters need to be examined by gathering MEG data from the same child, according to a machine learning-based classifier [26] is challenging. The lack of substantial, publicly available ASD datasets was the main drawback of the proposed machine-learning techniques [27]. ML neural models [28] have examined on limited number of kids as well and the algorithm only gave up when the child's hands created occlusion. Research on bio-inspired fuzzy optimizers [29] is required to improve the proposed algorithm dimensionality reduction process. Multikernel SVM [30] showed higher accuracy, nonetheless, the motion artifacts can further degrade the quality of physiological signals, leading to inaccurate measurements.

However, the modified Kalman filter [31] is still required to investigate system performance in naturalistic settings. Moreover, the GAN [32] is required to investigate the alternative architectures to SAGAN, such as Dense-Attentive GAN, for improving performance. The computational complexity and memory demands associated with VGG19 [33] can pose challenges to autism detection, while ResNet50 [34] may be susceptible to overfitting when dealing with limited data, which significantly affects the detection performance. Also, the 1D-CNN technique heavily depends on the purity and availability of the input data which might affect the performance [35]. Also, the multiple CNN model increases the complexity and is prone to overfitting when dealing with limited data [36]. Moreover, the MADDHM model, a novel approach, aims to address the aforementioned challenges associated with existing methods and enhance the effectiveness of autism detection.

3. Overview of the proposed multimodal autism detection process

It might be difficult to diagnose ASD in children because of its complexity and heterogeneity. At the moment, the majority of techniques in use primarily rely on one modality, provide little information, and frequently fall short of expectations. It remains difficult to investigate hidden correlations complementarily directly from the original data due to the variability of neurophysiologic and behavioral modalities. Multimodal fusion is an excellent solution to this problem. Multimodal fusion has gained a lot of attention lately, particularly in the field of medicine where it is being used to diagnose various illnesses like ASD. Therefore, multimodal autism detection is performed in our work using face images and EEG signals. The MADDHM scheme considers two modalities Face image and EEG signal.

- The proposed autism detection framework includes the following phases:
 - (i) Preprocessing
 - (ii) Feature Extraction
 - (iii) Feature level Fusion
 - (iv) Detection
- Preprocessing- Here, the two input modalities will be preprocessed. During the pre-processing stage, the input EEG signal is pre-processed by Wiener filtering and the face image is pre-processed using Gabor filtering.
- Feature Extraction
 - From the EEG signal data, features like Improved Singular Spectrum Entropy, correlation dimension, CSP based features will be extracted.

- From the face image, features will be extracted like the Improved Active Appearance model, Improved SLBT, and GLCM features as well.
- After the feature extraction, they are fused to form the combined feature set via the Improved Feature fusion process.
- Based on the fused features, the detection process will be performed via the hybrid model is the combination of CNN and Bi-GRU.
- Fig. 1 illustrates the developed methodology.

3.1. Preprocessing, feature extraction, and improved feature level fusion of developed model

In this research, the preprocessing, feature extraction and improved feature-level fusion of the developed model are described as follows. Let us consider the input EEG signal Φ_{EEG} and face image Φ_{IM} . Moreover, during the preprocessing stage, the input EEG signal Φ_{EEG} is pre-processed by Wiener filtering and the face image Φ_{IM} is preprocessed using Gabor filtering.

3.1.1. Preprocessing

When utilizing machine learning algorithms to identify autism, data pre-processing is essential. An essential step in the pre-processing of the data is outlier detection, which helps identify and handle data points that significantly deviate from the rest of the dataset. The EEG signal Φ_{EEG} is preprocessed via Wiener filtering. The face image Φ_{IM} is pre-processed using Gabor filtering

Wiener filtering:

The Wiener filter is a basic de-blurring filter used for de-noising. This is not the Wiener filter, which is widely used in image-reconstruction difficulties, but rather a basic local-mean filter [<https://docs.scipy.org/doc/scipy/tutorial/signal.html#wiener-filter>]. Wiener filters reduce the predicted value of the squared error signal, which is one of the key benefits of employing them for DSP noise reduction. The Wiener filter enhances the overall quality of the EEG signals by increasing the signal-to-noise ratio, which makes it easier to identify the relevant brain activity patterns associated with autism. So, the wiener filtering is a solid choice for noise reduction in EEG data which leads to generating effective preprocessed data for an efficient subsequent analysis. They are also optimal in the mean square error sense. If Φ_{EEG} is the input signal, the output.

$$\Phi_{WF} = \begin{cases} \frac{\sigma^2}{\sigma_{\Phi_{EEG}}^2} M_{\Phi_{EEG}} + \left(1 - \frac{\sigma^2}{\sigma_{\Phi_{EEG}}^2}\right) \Phi_{EEG}; & \sigma_{\Phi_{EEG}}^2 \geq \sigma^2 \\ M_{\Phi_{EEG}}; & \sigma_{\Phi_{EEG}}^2 < \sigma^2 \end{cases} \quad (1)$$

Where, $M_{\Phi_{EEG}}$ → local estimate of the mean, and $\sigma_{\Phi_{EEG}}^2$ → local estimate of the variance. The parameter σ^2 is a threshold noise parameter. Here, the preprocessed EEG signal is labelled as Φ_{WF} .

Gabor filtering: Gabor filters [37] are bio-inspired convolutional kernels that have a wide range of applications in image processing and computer vision. The benefit of employing Gabor filters seems to lie in their capacity to offer a certain level of invariance to translation, intensity, and orientation. In autism detection, Gabor filtering is effective for texture analysis and edge detection, which is useful for preprocessing facial images. Moreover, the local spatial frequency information and orientation, are important features for characterizing facial textures so the Gabor filter is the best choice, which captures both fine and coarse texture details in facial images as well as increases the facial image quality for further effective analysis. These filters have two intriguing properties: orientation selectivity and location frequency, which are akin to the human visual system. Thus, image analysis utilizing these filters was discovered to be very beneficial for texture classification and representation. Image processing applications including optical character recognition, iris recognition, and fingerprint recognition greatly

benefit from the utilization of the Gabor space. A 2-D Gabor filter in the spatial domain is a Gaussian kernel technique produced by a complex sinusoidal plane wave, as stated in Eq. (2).

$$\left. \begin{aligned} \Phi_{GF} &= \frac{f^2}{\pi\gamma\eta} \exp\left(-\frac{\chi^2 u^2 + \delta^2 v^2}{2SD^2}\right) \exp(j2\pi f u' + \varphi) \\ u' &= u\cos\theta + v\sin\theta \\ v' &= -u\sin\theta + v\cos\theta \end{aligned} \right] \quad (2)$$

where \rightarrow sine frequency, $\theta \rightarrow$ orientation of the Gabor method, $\varphi \rightarrow$ phase offset, χ and $\delta \rightarrow$ Gaussian sharpness with the major axis parallel and the minor axis perpendicular to the wave, respectively. σ standard deviation of the Gaussian envelope, and $\gamma\eta \rightarrow$ spatial aspect ratio. Here, the preprocessed face image is labelled as Φ_{GF} .

3.1.2. Feature extraction

Feature extraction is a method of transforming raw data into numerical features that can be managed while preserving the details included in the initial set of data. From the filtered signal data Φ_{WF} , the features like improved singular spectrum entropy-based features, correlation dimension, and common spatial patterns are extracted. From the filtered face image Φ_{GF} , features like improved AAM, improved SLBT, and GLCM features are extracted.

For EEG-filtered signal data:

ISSD-based feature extraction: The Improved Singular Spectrum Decomposition (ISSD) approach is suggested in this part to address the difficulty of the original SSD in determining the embedding [38] dimension according to the empirical principle. The ISSD can lead to better separation of signal components from the data, even in the presence of complex interference. This can help in extracting meaningful information from the preprocessed EEG signal and improving the overall decomposition quality. The ISSD approach is summarized for a given time series $y(i)$ as follows:

- (1) Create a new trajectory matrix: With the original SSD, a new trajectory matrix Y is first constructed using the time series $y(i) = \{1, 2, 3, 4, 5\}$ with embedding dimension $N = 3$ as an example

$$Y = \left[\begin{array}{cc|ccc} & & 1 & & & \\ & & 1 & 2 & & \\ 1 & 2 & 3 & 4 & 5 & \\ 2 & 3 & 4 & 5 & * & \\ 3 & 4 & 5 & * & * & \end{array} \right] \quad (3)$$

- (2) To find the embedding dimension N adaptively, use the Cao technique. First, for various embedding dimensions, the nearest neighbor point's distance variation in the phase space is determined

$$d(j, N) = \frac{\sqrt{\sum_{j=1}^i (|Y_j(N+1)| - |Y_{i(j,N)}(N+1)|)^2}}{\min\left(\sqrt{\sum_{j=1}^i (Y_j(N))^2}, \sqrt{\sum_{j=1}^i (Y_{i(j,N)}(N))^2} + \beta\right)} \quad (4)$$

where $j = 1, 2, 3, \dots, K - N$, $\beta = 0.6 Y_j(N+1) \rightarrow j^{th}$ phase space reconstruction signal, $(N+1)$ the embedding dimension, $Y_{i(j,N)}(N+1) \rightarrow$ phase space reconstruction signal nearest to $Y_j(N+1)$, $i(j, N) \rightarrow$ an integer and satisfies $K - N < i(j, N) \leq K - N$. Also, contra harmonic mean is used for distance variation, which can increase the accuracy of distance variation computation, leading to more reliable decomposition results.

$$E_1(l) = \sum_{j=1}^{K-l} \frac{d^2(j, N)}{d(j, N)} \quad (5)$$

where $E_1(l)$ is the average of all $d(j, N)$. Finally, the changing behavior of $E_1(l)$ is checked by

$$E_2(l) = E_1(l+1) / E_1(l) \quad (6)$$

In Eq. (6), if $N > N_0$, $E_1(l)$ achieves stability and has less fluctuation, the minimum embedding dimension is automatically determined as $N_0 + 1$.

(3) The m^{th} sub-series is reconstructed orderly in different frequency bands.

(4) Place the stopping condition owing to Pearson correlation analysis. Specifically, at the m^{th} iterations, the resulting series $v^{(m+1)}(i)$ is obtained by subtracting the m^{th} sub-series $g^{(m)}(i)$ from the original series, that is, $v^{(m+1)}(i) = v^{(m)}(i) - g^{(m)}(i)$. Then, the resulting series $v^{(m+1)}(i)$ is regarded as the input series to continue the next $m + 1$ iteration. Finally, cross-correlation coefficient between the residual series $v^{(j)}(i)$ and the original series $y(i)$ is calculated by

$$\rho(v^{(j)}, y) = \left| \frac{E(v^{(j)}(i) - \mu_j) - E(v^{(j)}(i) - \mu_y)}{SD_j SD_y} \right| \quad (7)$$

where μ_j and μ_y is mean value of $v^{(j)}(i)$ and $y(i)$, σ_j and σ_y is the standard deviation of $v^{(j)}(i)$ and $y(i)$. For Pearson correlation analysis, we define one threshold as $\epsilon = 0.1 \times \max(\rho(v^{(j)}, y))$. The original time series $y(i)$ is decomposed into several sub-series and the residual $v^{(l+1)}(i)$.

$$y(i) = \sum_{h=1}^l g^{(h)}(i) + v^{(l+1)}(i) \quad (8)$$

where l is the count of sub-series and $g^{(h)}(i)$ is the h^{th} sub-series. The ISSD's output is denoted as Φ_{ISSD} .

Correlation Dimension-based feature extraction: The correlation integral is approximated by the correlation dimension [39] and represented as

$$\Phi_{Dim} = \lim_{r \rightarrow 0} \frac{\log(C(r))}{\log(r)} \quad (9)$$

$$C(r) = \frac{1}{D(D-1)} \sum_p \sum_q V(r, |W_p - W_q|) \quad (p \neq q) \quad (10)$$

Here, r denotes the (radius). The output of the Correlation Dimension is denoted as Φ_{Dim} .

CSP-based feature extraction: The CSP [40] realizes the best spatial filter $\omega \in \Phi_{GF}^{M \times 1}$, it reduces the ratio between the two activities' variances. Eq. (10) expresses this objective function.

$$I_1(X, \omega) = \frac{\frac{1}{n_1} \sum_{O \in j_1} \| \omega^\top X_o^{task} \|^2}{\frac{1}{n_2} \sum_{O \in j_2} \| \omega^\top X_o^{task} \|^2} = \frac{\frac{1}{n_1} \sum_{O \in j_1} \omega^\top G_o \omega}{\frac{1}{n_2} \sum_{O \in j_2} \omega^\top G_o \omega} = \frac{\omega^\top \bar{G}_1 \omega}{\omega^\top \bar{G}_2 \omega} \quad (11)$$

Where, $G_o \rightarrow$ autocorrelation matrix of X_o^{task} and $n_1, n_2 \rightarrow$ amount of each type of job. Because of the frequent fluctuations and high temporal resolution of EEG, the best filter is used to combine all of the channels form a single series of data from every trial, and variance is derived as a crucial feature. Covariance in convolution is determined for every class using Eqs. (12) and (13).

$$SI_1 = \text{cov}(X_1) \quad (12)$$

Table 2
GLCM features.

S. No.	Features	Formula
1.	“Entropy	$Entropy = - \sum_a \sum_b Q_{ab} \log_2 Q_{ab}$
2.	Energy	$Energy = \sum_a \sum_b Q_{ab}^2$ here Q_{ab} is the $(a, b)^{th}$ entry in GLCM
3.	Variance	$Var = \sum_a \sum_b (a - \bar{R})^2 Q_{ab}$, where \bar{R} specifies the mean of Q_{ab}
4.	Contrast	$Con = \sum_a \sum_b (a - b)^2 Q_{ab}$
5.	Correlation	$Cor = \frac{\sum_a \sum_b (ab) Q_{ab} - \bar{R}_a \bar{R}_b}{SD_a SD_b}$, where $SD_a, SD_b, \bar{R}_a, \bar{R}_b$ are the std deviations and mean of Q_{ab}, Q_b
6.	Homogeneity	$H = \sum_a \sum_b \frac{1}{1 + (a - b)^2} Q_{ab}$
7.	Sum Variance	$SV = \sum_{a=2}^{20q} (a - SE)^2 Q_{a+b}(a)$
8.	Sum Entropy	$SE = \sum_{a=2}^{20q} Q_{a+b}(a) \log\{Q_{a+b}(a)\}$
9.	Difference Entropy	$DE = \sum_{a=0}^{O_{q-1}} Q_{a-b}(a) \log\{s_{a-b}(a)\}$

$$SI_2 = \text{cov}(X_2) \quad (13)$$

The developed CSP-based feature is indicated as Φ_{CSP} .

Finally, the extracted EEG signal features $\Phi_{FE}(EEG)$ are determined in Eq. (14).

$$\Phi_{FE}(EEG) = [\Phi_{ISD} \quad \Phi_{Dim} \quad \Phi_{CSP}] \quad (14)$$

For the filtered face image:

Improved AAM (IAAM) based feature extraction:

AAM [41]: In order to find the edge intensity, the IAAM is used. From the ASM model, the AAM model developed. Triangles, whose vertices act as ASM marking points, divide the face into smaller portions. Before restarting PCA analysis, the active appearance model links the shape and texture PCA components for a single class of vector with a weight vector.

A metric that combines edge intensity and texture information is utilized to characterize the image structure. The low-frequency coefficient of each image can be obtained by applying the translation-invariant wavelet transform to a facial image from the training set. The edge intensity is then combined with the low-frequency coefficients. Assume $c(x, z)$ denotes the low-frequency coefficients obtained via translation invariant wavelet transform, the edge intensity $e(x, z)$ can be computed as below:

$$e(x, z) = (c(x, z)^* J_1)^2(x, z) + (c(x, z)^* J_2)^2(x, z) + (c(x, z)^* J_3)^2(x, z) \quad (15)$$

As per the improved AAM approach, the J value is calculated as per Eq. (16), which enables more accurate localization of facial landmarks, which are crucial for aligning facial images within the AAM framework. Additionally, the improved AAM model is useful for significantly capturing facial shape and texture information which improves the precision of feature extraction and facilitates accurate representation of facial variations.

$$J(x) = \left| \frac{\{(A_x - A_x^L) + (A_x - A_x^R)\}}{(A_x - A_x^T) + (A_x - A_x^B)} - \frac{\frac{1}{2}(A_{cx} - A_x^L) + (A_{cx} - A_x^R)}{(A_{cx} - A_x^T) + (A_{cx} - A_x^B)} \right| \quad (16)$$

Here, A_x is pixel values, L and R represent the pixel's left and right values, A_{cx} denotes the pixel's central value, and $x = 1, 2, \dots, k$. IAAM feature is denoted as Φ_{IAAM} .

GLCM-based feature extraction: The spatial relationship between pixels is calculated using it. With GLCM, textural properties are retrieved. GLCM [42] is a second-order statistics approach based on (local) gray-level information in a pair of pixels. Table 2 has a

description of GLCM.

The extracted GLCM features are denoted by Φ_{GLCM} .

Proposed SLBT-based feature extraction: SLBT [43] functions combine shape and texture information. In order to find the gradient value of each image pixel, the proposed SLBT is used. The proposed SLBT generates robust representations of facial expressions by learning a set of discriminative templates from the pre-processed image sequences and also captures subtle differences in facial expressions that cannot be recognised by conventional methods.

Let $\Phi_{GF} = [\Phi_{GF1}, \Phi_{GF2}, \dots, \Phi_{GFp}]$ represent M training set images with $Z = [Z_1, Z_2, \dots, Z_p]$ as its shape landmark points. In the training set, every shape vector z may be represented as in Eq. (17).

$$Z \approx \bar{Z} + \theta_k \hat{A}_k, \quad (17)$$

$$\hat{A}_k = \theta_k^T (Z - \bar{Z})$$

Where, $\theta_k \rightarrow$ eigenvectors of largest eigenvalues, $\bar{Z} \rightarrow$ mean shape, and $\hat{A}_k \rightarrow$ shape model or weights parameters. Consider a 3x3 window with center pixel (\hat{u}_c, \hat{v}_c) intensity value be g_c and local texture as $\hat{Q} = \hat{q}(g_i)$ where $g_i \in \{0, 1, 2, 3, 4, 5, 6, 7\}$ corresponds to the grey values of 8 surrounding pixels. The LBP pattern at the center pixel g_c can be obtained using Eq. (19). The function $\hat{a}(\hat{n})$ is defined as,

$$\hat{a}(\hat{n}) = \begin{cases} 1, & \hat{n} > 0 \\ 0, & \hat{n} \leq 0 \end{cases} \quad (18)$$

$$LBP(\hat{u}_c, \hat{v}_c) = \sum_{i=0}^7 \hat{a}(g_i - g_c) 2^i \quad (19)$$

As per the proposed logic,

$$g_i = \hat{x} - \frac{1}{|\hat{x}|} adj(\hat{x}) \quad (20)$$

$$g_c = \sum_{i=1}^n \frac{\hat{n}}{g_i} \quad (21)$$

Here, g_i is the gradient value of adjacent pixels, and g_c is the harmonic mean of the gradient value for adjacent pixels.

Let $U = [U_1, U_2, \dots, U_p]$. The texture modelling is performed via PCA as in Eq. (22)

$$\hat{A}_i = \theta_i^T (U - \bar{U}) \quad (22)$$

Eq. (23) produces the total shape and texture parameter vector. The shape texture parameter regulating texture, global and local shape may be generated by performing PCA on the combined parameter vector as stated in Eq. (24).

$$\hat{A}_{ki} = \begin{pmatrix} w_k \hat{A}_k \\ \hat{A}_i \end{pmatrix} \quad (23)$$

$$\hat{C} = \theta_k^i (\hat{A}_{ki} - \overline{\hat{A}_{ki}}) \quad (24)$$

where $\hat{C} \rightarrow$ shape texture parameter, w_k represents the weight diagonal matrix of every shape parameter, $\theta_k^i \rightarrow$ eigenvectors and $\overline{\hat{A}_{ki}}$ \rightarrow mean vector. The obtaining SLBT feature is denoted as Φ_{PSLTB} .

Moreover, the entire face image extracted features $\Phi_{FE}(IM)$ are determined in Eq. (25).

$$\Phi_{FE}(IM) = [\Phi_{IAAM} \quad \Phi_{GLCM} \quad \Phi_{PSLTB}] \quad (25)$$

3.1.3. Improved feature level fusion

After feature extraction, they are fused to form the combined feature set by an improved feature-level fusion process. Here, both the face image-based features and EEG signal-based features are fused together.

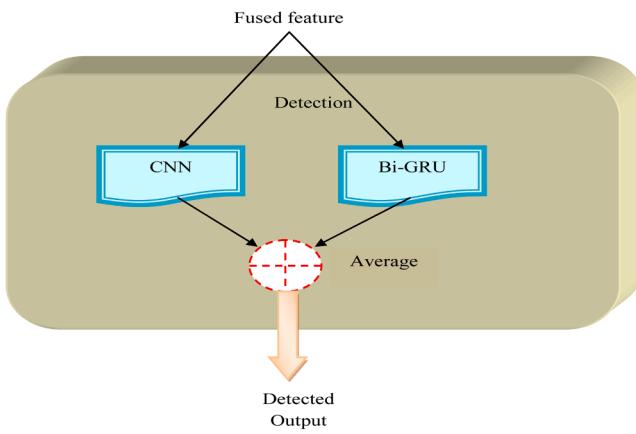


Fig. 2. Hybrid classifier model.

The improved feature fusion process remove the impact of outliers leading to obtaining more reliable fused features for a further effective diagnosis process.

$$\hat{Z} = TM_{\Phi_{FE}(EEG)} * (\Phi_{FE}(EEG) \otimes (1 - \hat{R})). \quad (26)$$

$$TM_{\Phi_{FE}(IM)} * (\Phi_{FE}(IM) \otimes \hat{R})$$

Where, $TM_{\Phi_{FE}(EEG)}$ is the trimmed mean of $\Phi_{FE}(EEG)$, and $TM_{\Phi_{FE}(IM)}$ is the trimmed mean of $\Phi_{FE}(IM)$.

$$TM = \frac{1}{\hat{o} - 2\hat{K}} \sum_{i=\hat{K}+1}^{\hat{o}-\hat{K}} f(\Phi_{FE(i)}) \quad (27)$$

3.2. Diagnosis process via hybrid classifier and its objective function

The fused features from the previous step are subjected to a hybrid classification model, where autism is detected. A hybrid classifier combining CNN and Bi-GRU classifiers is used in this step. CNNs may not generalize well, leading to suboptimal performance when dealing with limited data. It can inadvertently learn biases present in the training data. If the training data is biased, the model may produce biased predictions. However, the CNN has dominant advantageous characteristics which can effectively capture intricate patterns and textures, and provide rich, high-level features that can enhance the model's ability to detect subtle differences associated with autism. By leveraging the strengths of individual models make prior to improve the detection efficiency. The BiGRU and CNNs often learn complementary features from the input. While BiGRU focuses on capturing dependencies and patterns over time, CNNs excel at capturing local spatial patterns. By combining the outputs of both models, the final classifier can benefit from a richer feature representation, potentially leading to improved detection performance. Thereby, the output of both classifiers is averaged to obtain the final detected output. Here, \hat{z} is given as the input of the classification phase for CNN and Bi-GRU classifiers. Fig. 2 illustrates the hybrid classifier model.

3.2.1. Bi-GRU

In this case, \hat{Z} is used as an input to Bi-GRUs [44]. It is advantageous to evaluate sequential data in a uniform format. RNNs are also very excellent at encoding sequential data. The two processes that comprise the Bi-GRU's computing process are forward and reverse sequence information transfers. The forward GRU for a certain segment is determined using the following formula: where, $\hat{E} = (\hat{e}_1, \hat{e}_2, \dots, \hat{e}_j)$. The following shows how the forward GRU is calculated.

$$\hat{D} = \mathbb{N}(\hat{w}_{\hat{e}\hat{D}}\hat{Z} + \hat{w}_{\hat{G}\hat{D}}\hat{G}_{\hat{r}-1} + \hat{B}_{\hat{D}}) \quad (28)$$

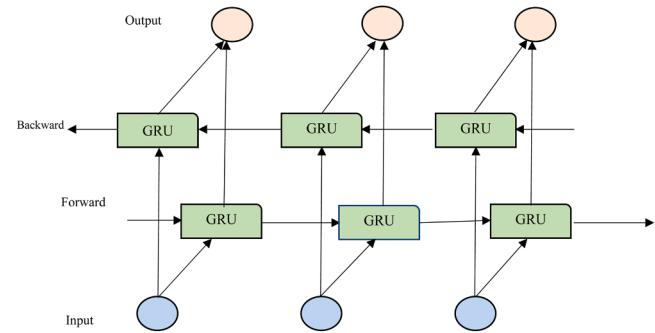


Fig. 3. Structure of BiGRU.

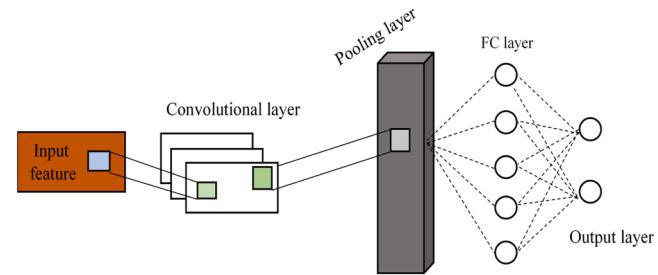


Fig. 4. Architecture of CNN.

$$\hat{P} = \mathbb{N}(\hat{w}_{\hat{e}\hat{P}}\hat{Z} + \hat{w}_{\hat{G}\hat{P}}\hat{G}_{\hat{r}-1} + \hat{B}_{\hat{P}}) \quad (29)$$

$$H = \tanh(\hat{w}_{\hat{e}H}\hat{Z} + \hat{w}_{\hat{G}H}(\hat{D}\Theta\hat{G}_{\hat{r}-1}) + \hat{B}_H) \quad (30)$$

$$\hat{G} = (1 - \hat{P})\Theta\hat{G}_{\hat{r}-1} + \hat{P}\Theta H \quad (31)$$

Here, \hat{Z} and \hat{w} → bias vector & weight matrix, accordingly. \mathbb{N} → sigmoid function, $\hat{G}_{\hat{r}}$ → hidden state of the current time step \hat{r} , and Θ → element-wise multiplication. The results of the backward and forward GRUs → $\hat{G}_{\hat{D}}$ & $\hat{G}_{\hat{D}}$. Fig. 3 shows the structure of BiGRU network. Moreover, the Bi-GRU outcomes are determined as $\hat{G}_{\hat{D}}^{Bi-GRU} = [\vec{G}_{\hat{D}}; \overleftarrow{G}_{\hat{D}}]$.

3.2.2. CNN

CNN is given the extracted features \hat{z} as a source. The visual system of living organisms is modelled by CNN [45], a unique multilayer NN or DL architecture. Furthermore, CNN a well-known deep learning architecture is made up of fully connected, convolutional, and pooling layers. Fig. 4 illustrates the architecture of CNN.

Convolutional Layer: The feature values are assessed using Eq. (32) at the provided position (\hat{x}, \hat{z}) in the S^{th} layer of the appropriate \hat{b}^{th} feature map.

$$F_{\hat{x}, \hat{z}, \hat{b}}^S = \tilde{W}_b^{S\hat{r}}\hat{Z} + \hat{l}_b^S \quad (32)$$

Where, \tilde{W}_b^S → weight used in training & \hat{l}_b^S → bias term of the S^{th} layer-specific \hat{b}^{th} filter, \hat{Z} → the patched input in the position (\hat{x}, \hat{z}) of S^{th} layer.

Dropout layer: CNN has a Dropout layer as well. The Dropout layer keeping all other neurons operational serves as a mask while eliminating some neurons' contributions to the succeeding layer.

Pooling layer: The subsampling of the feature map is the pooling layer's primary function. The convolutional operations are followed to create these maps. Simultaneously, it preserves most of the prominent characteristics during the entire pooling phase.

Flatten layer: A flattened layer condenses the spatial dimensions of an

Table 3
Hyper-parameter of classifiers.

Hyper-parameters		
Bi_GRU		
activation	sigmoid	
loss	sparse_categorical_crossentropy	
optimizer	rmsprop	
CNN		
activation	softmax	
loss	categorical_crossentropy	
optimizer	Adam	
DBN		
activation_function	Sigmoid	
DNN		
activation	Sigmoid	
loss	sparse_categorical_crossentropy	
optimizer	Rmsprop	
LSTM		
activation	sigmoid	
loss	sparse_categorical_crossentropy	
optimizer	sgd	

input to its channel dimension. Eq. (33) determines the loss of CNN.

$$Loss = \frac{1}{m} \sum_{n=1}^m \tilde{k}(k; \tilde{V}^{(n)}, \tilde{O}^{(n)}) \quad (33)$$

The input-output relations are $\tilde{O}^{(n)}$, $\tilde{V}^{(n)}$ and $\tilde{Z}^{(n)}$ display the CNN output, related target labels, and input feature. The CNN output is displayed as $\tilde{O}^{(n)}$. The detection outcome O^* is specified in Eq. (34).

$$O^* = \frac{\tilde{G}_D^{Bi-GRU} + \tilde{O}^{(n)}}{2} \quad (34)$$

3.2.3. Objective function

The objective function is described in Eq. (35), which describes the loss LS among the actual & predicted values by taking the average of CNN loss and Bi-GRU loss. Table 3 represents the hyper-parameters of classifiers.

$$Obj = \min(LS) \quad (35)$$

4. Results and discussions

4.1. Simulation setup

The MADDHM model was implemented in the PYTHON tool, and its findings were examined. The datasets were obtained from [<https://www.kaggle.com/datasets/disbeat/bciaut-p300>, <https://www.kaggle.com/general/123978>]. Based on negative, positive, and other metrics, the MADDHM model for diagnosing autism was determined by varying the learning percentages to 60, 70, 80, and 90. The comparison was made over the existing models such as Bi-GRU, CNN, DBN [46], DNN [47], LSTM, and SVM and the state-of-the-art models like VGG19 [33] and ResNet50 [34] respectively.

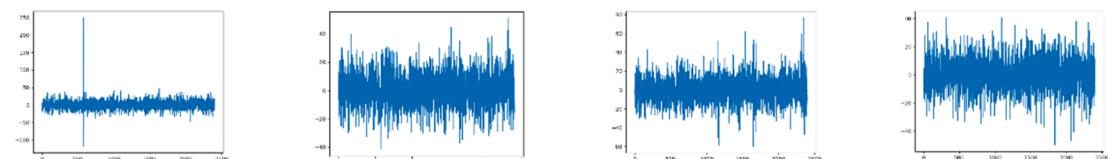
4.2. Dataset description

Multi-Session and Multi-Subject Benchmark Dataset on Autism for P300-BCI (BCIAUT_P300) (Dataset1): This dataset represents the complete EEG recordings of a feasibility clinical trial (clinical-trial ID: NCT02445625 — clinicaltrials.gov) that tested a P300-based Brain Computer Interface to train youngsters with ASD to follow social cues. A total of 15 ASD individuals underwent 7 sessions of P300-based BCI joint-attention training, for a total of 105 sessions. The dataset was used for the 2019 IFMBE Scientific Challenge organized during MEDICON 2019 where, in two phases, teams from all over the world tried to achieve the best possible object-detection accuracy based on the P300 signals.

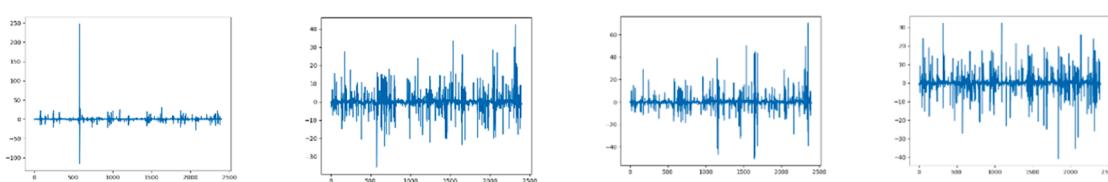
Autistic Children Facial Image Data Set (Dataset2): A data set of facial images of children with Autism and facial images of children without Autism. The data set is suitable for training a classifier. There are 1327 images of Autistic children's faces and for balance 1327 images of non-autistic children in the training set. The test set contains 140 images of autistic children's facial images and 140 images of non-autistic children.

The features present in the dataset are fused together considering two modalities Face and EEG. Assume, that the Face and EEG belong to the same person.

Sample data



Wiener filter-based Preprocessing



Improved singular spectrum feature-based outcome

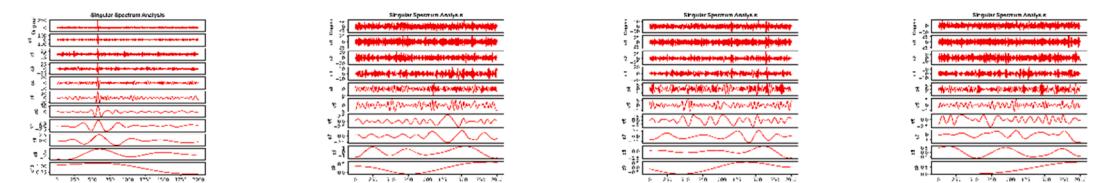


Fig. 5. Analysis of processing outcomes of the EEG signal.

Sample image

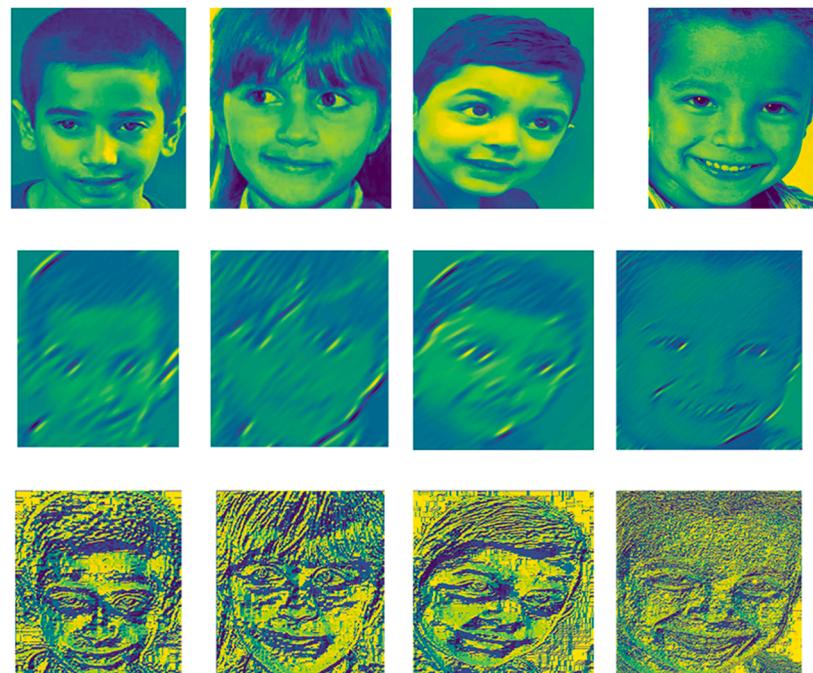


Fig. 6. Analysis of processing outcomes of the facial images.

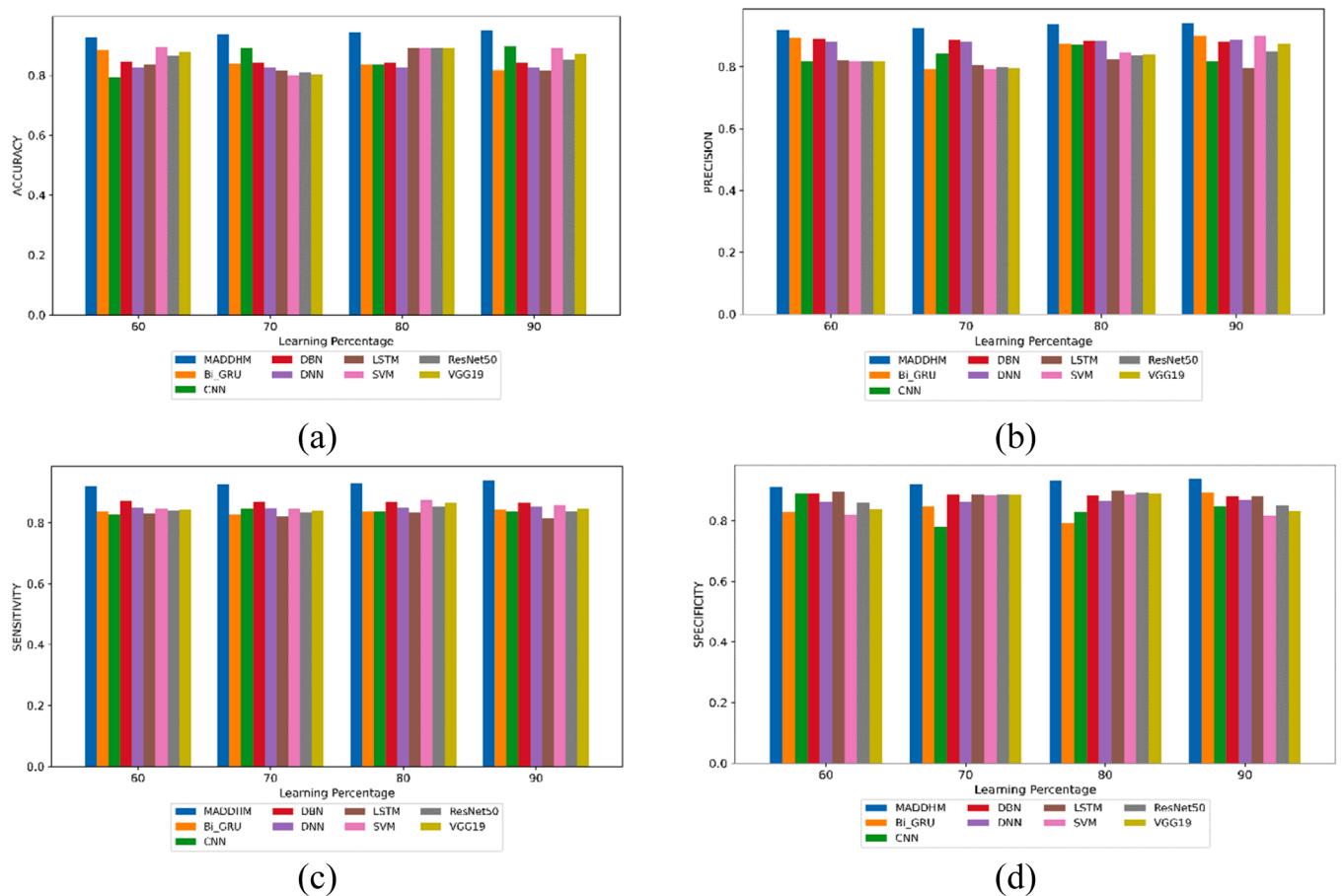


Fig. 7. Positive measure analysis on MADDHM & conventional models for dataset1 a) Accuracy b) Precision c) Sensitivity d) Specificity.

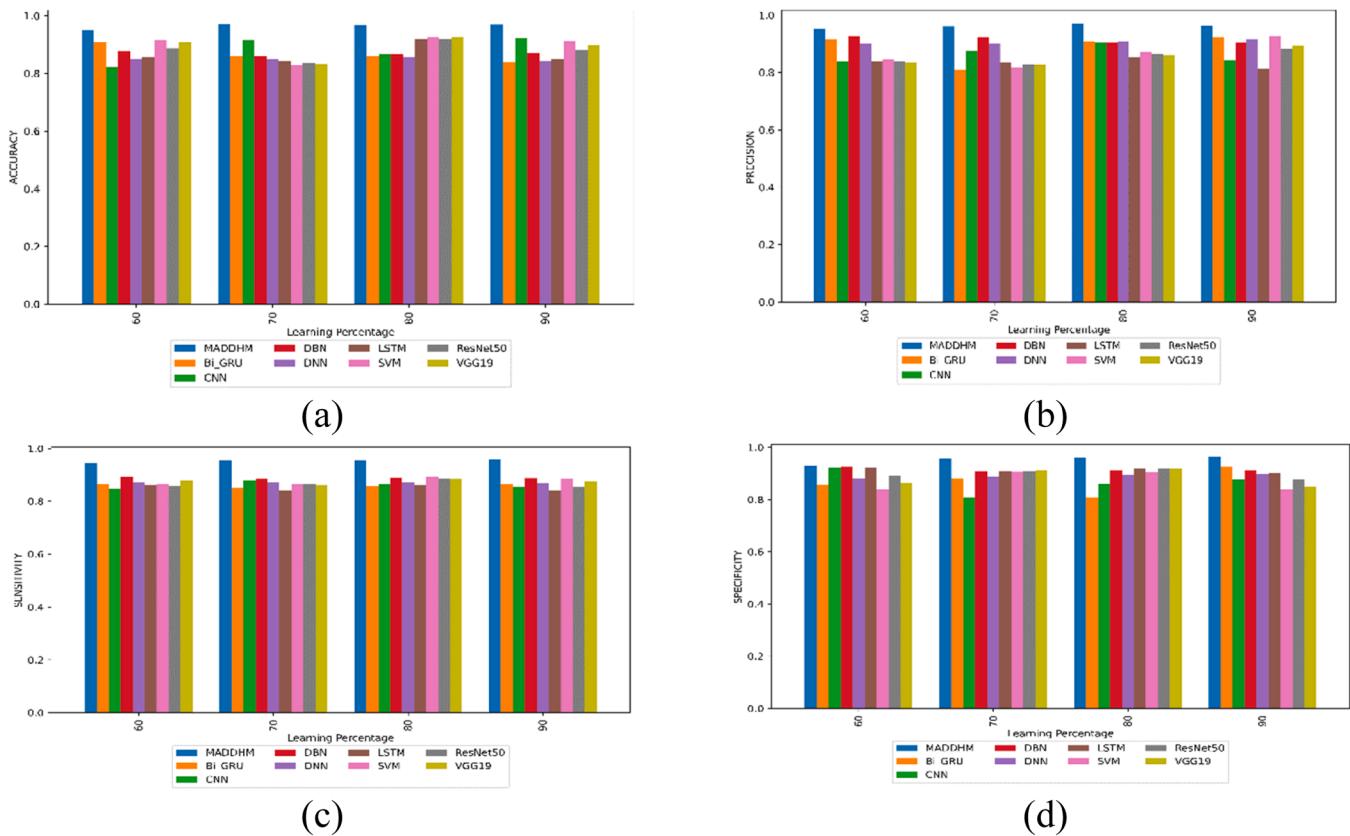


Fig. 8. Positive measure analysis on MADDHM & conventional models for dataset2 a) Accuracy b) Precision c) Sensitivity d) Specificity.

4.3. Analysis of processing results for EEG signal and facial image

Figs. 5 and 6 illustrate the sample data, the outcomes of pre-processing, and features extraction processes of the two modalities such as EEG signal and facial image respectively.

4.4. Analysis of MADDHM over existing models for detecting autism on positive measures

Figs. 7 and 8 depict the performance of the MADDHM model for detecting autism in comparison to existing techniques based on positive criteria such as Accuracy, Precision, Specificity, and Sensitivity. Further, the MADDHM model attained the maximal accuracy (0.94) for detecting autism at a learning percentage of 90 %, than other traditional approaches including Bi-GRU, CNN, DBN [46], DNN [47], LSTM, VGG19 [33], ResNet50 [34] and SVM, respectively attains least accuracy value. Furthermore, compared to other standard schemes for detecting

multimodal autism at a learning percentage of 80 %, the MADDHM model holds maximum specificity (0.92). The MADDHM model offers more sensitivity (0.95) for detecting autism at a 90 % learning percentage than a learning percentage of 60 %. Additionally, the MADDHM model yields better precision results for detecting autism at a learning percentage of 70 % than other schemes such as Bi-GRU, CNN, DBN [46], DNN [47], LSTM, VGG19 [33], ResNet50 [34] and SVM, respectively. For dataset2, the MADDHM model achieved a maximum precision value about 0.95 at 90 percent of learning whilst BiGRU=0.86, CNN=0.783, VGG19=0.83, ResNet50=0.86, SVM=0.77. Thus, an improvement of the MADDHM model is attained for detecting autism more efficiently using improved features and a hybrid classifier concept.

4.5. Analysis of MADDHM over existing models for detecting autism on negative measures

Figs. 9 and 10 show the comparison of the MADDHM model for

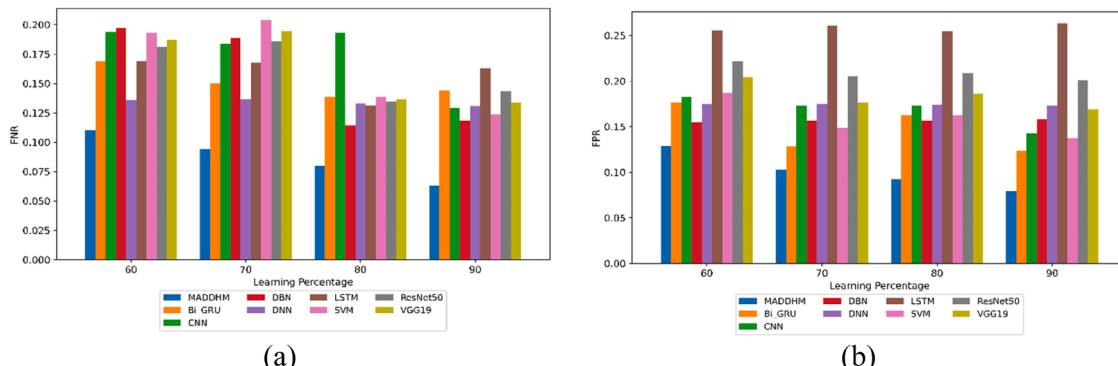


Fig. 9. Negative measure analysis on MADDHM model to other methods for dataset1 a) FNR b) FPR.

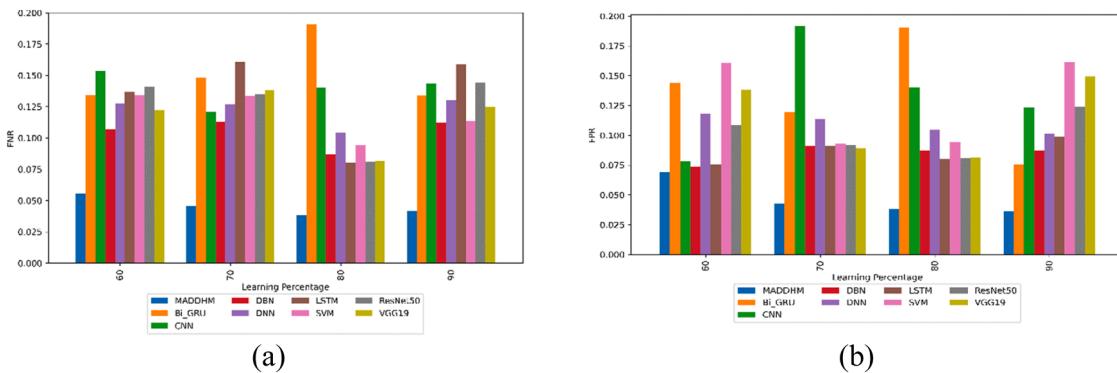


Fig. 10. Negative measure analysis on MADDHM model to other methods for dataset2 a) FNR b) FPR.

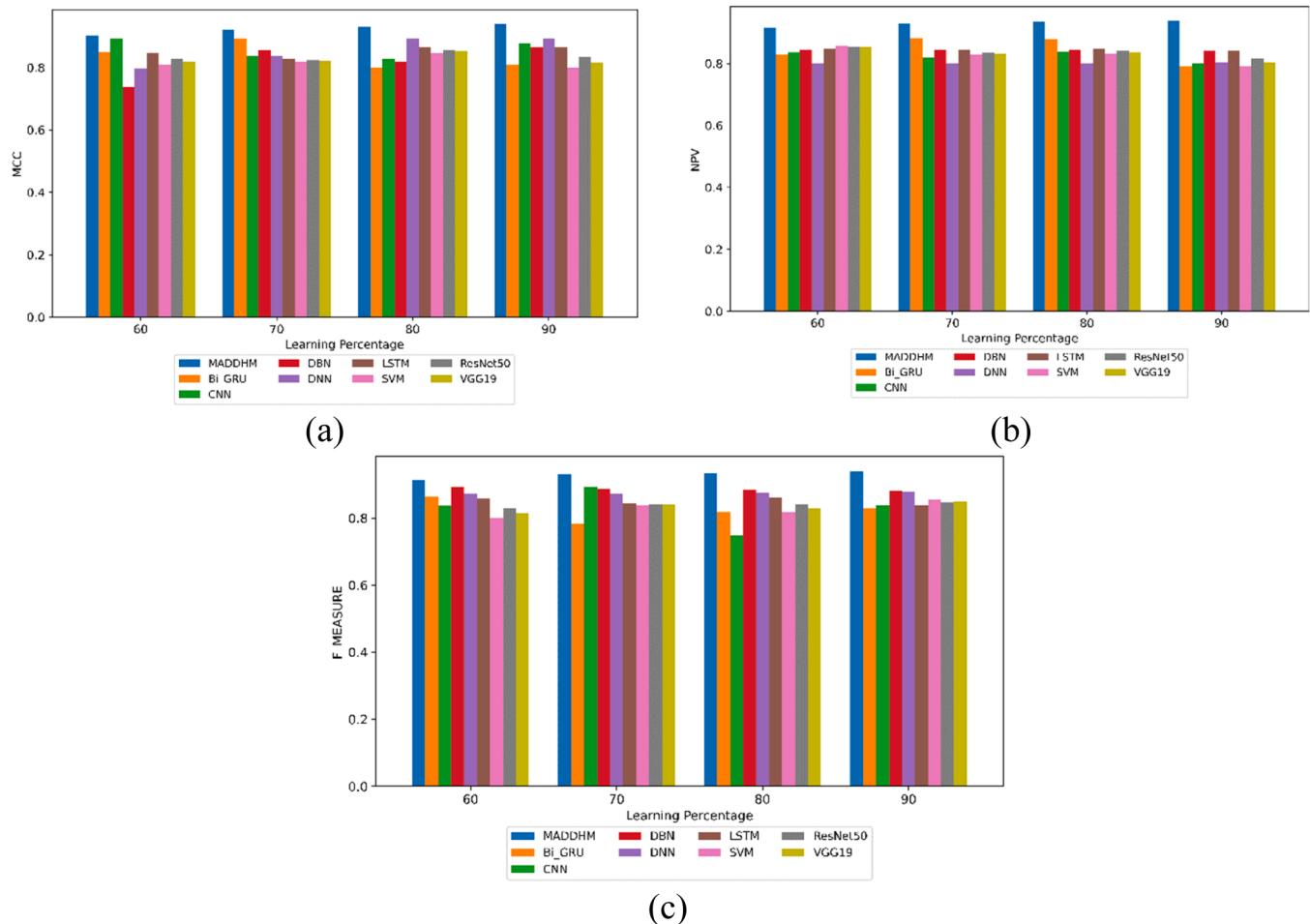


Fig. 11. Other measure analysis on the MADDHM model & previous schemes for dataset1 a) MCC b) NPV c) F-measure.

detecting autism to conventional schemes such as Bi-GRU, CNN, DBN [46], DNN [47], LSTM, VGG19 [33], ResNet50 [34] and SVM, respectively with regard to negative metrics (FNR, & FPR) for datasets 1 and 2. Further, on dataset 1, the MADDHM model has achieved fewer FPRs with higher performance (0.10) for detecting autism at a learning percentage of 70 % than other learning percentages of 60 %. Similarly, at a learning percentage of 90 %, the MADDHM model for detecting autism outperformed other schemes including Bi-GRU, CNN, DBN [46], DNN [47], VGG19 [33], ResNet50 [34], LSTM, and SVM, respectively in terms of minimal FNR values. Consequently, for dataset 2, the MADDHM model offers a lower FPR of around 0.67 at 90 % of learning meanwhile the existing models achieved greater FPR value than the MADDHM. The

lesser rate of FNR and FPR contributes to improving the detection performance of autism. This attainment of better performance ensures accurate diagnosis with less error rate. Here, the improvements in feature extraction play a major role as it extracts as much information to train the hybrid model.

4.6. Analysis of MADDHM over existing models for detecting autism on other measures

Figs. 11 and 12 represent the comparison of the MADDHM model over the conventional schemes including Bi-GRU, CNN, DBN [46], VGG19, ResNet50, DNN [47], LSTM, and SVM, respectively for

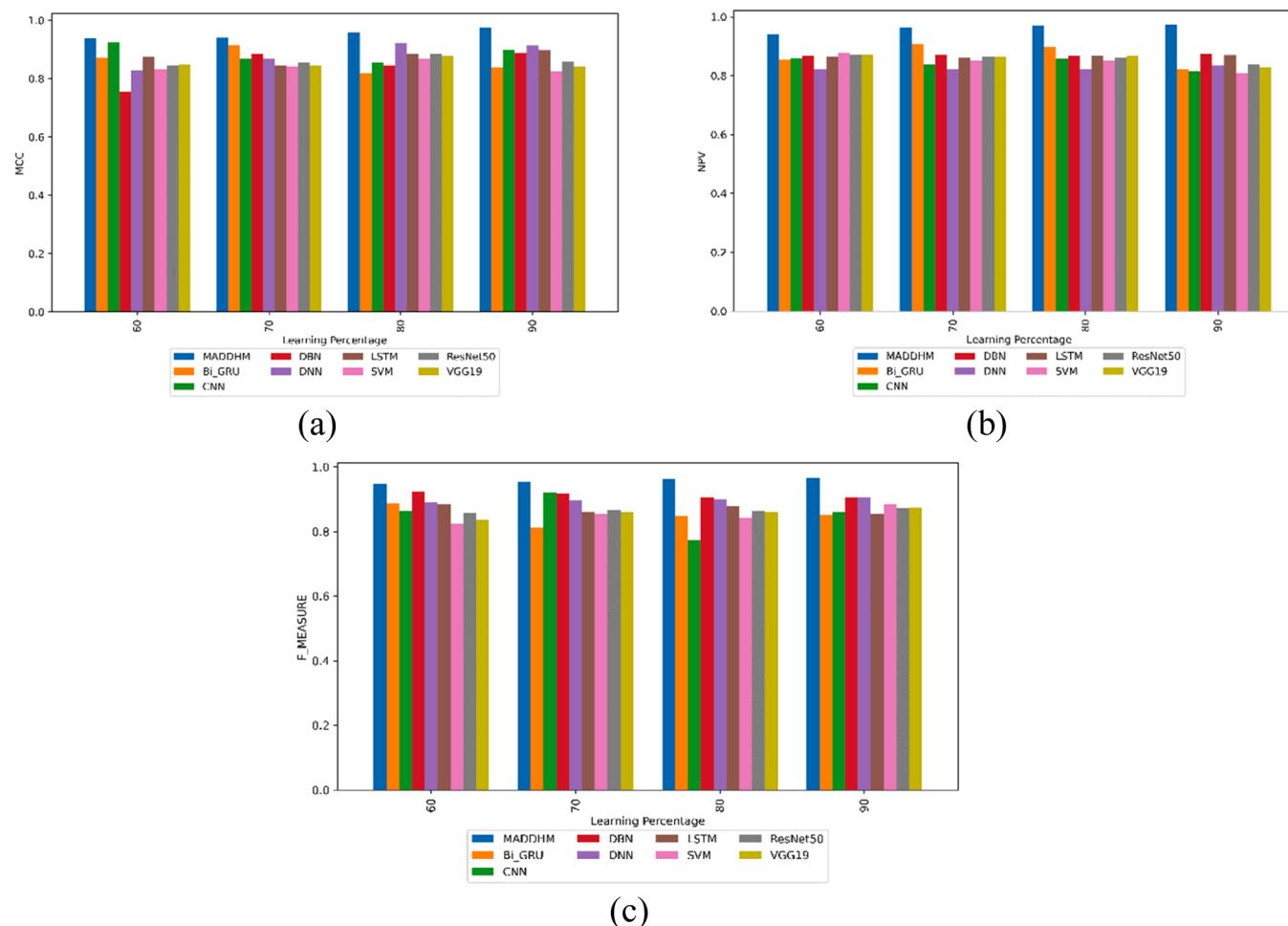


Fig. 12. Other measure analysis on the MADDHM model & previous schemes for dataset2 a) MCC b) NPV c) F-measure.

Table 4
Ablation analysis.

Metrics	Hybrid model without feature extraction	Hybrid model with conventional SSE	Hybrid model with conventional AAM feature	Hybrid model with conventional SLBT feature	Hybrid model with conventional feature level fusion	MADDHM scheme	Unimodal features for the face	unimodal features for EEG	Without preprocessing
Specificity	0.891	0.875	0.890	0.801	0.889	0.921	0.900	0.828	0.830
Sensitivity	0.836	0.847	0.892	0.827	0.847	0.927	0.833	0.892	0.903
Accuracy	0.802	0.817	0.827	0.847	0.817	0.935	0.879	0.874	0.910
Precision	0.789	0.800	0.799	0.821	0.829	0.924	0.719	0.820	0.850
F-measure	0.817	0.792	0.818	0.877	0.856	0.928	0.800	0.902	0.900
MCC	0.847	0.892	0.845	0.800	0.836	0.920	0.783	0.792	0.879
NPV	0.882	0.837	0.837	0.837	0.818	0.930	0.863	0.800	0.899
FPR	0.149	0.192	0.189	0.138	0.149	0.102	0.162	0.198	0.102
FNR	0.187	0.163	0.156	0.153	0.132	0.093	0.189	0.202	0.120

detecting autism in relation to other metrics (F-measure, MCC, and NPV) for datasets 1 and 2 respectively. While the learning rate is 70 %, the MADDHM model achieved a higher F-measure value (0.92) for detecting autism; however, the previous method like Bi-GRU, CNN, DBN [46], DNN [47], LSTM, VGG19 [33], ResNet50 [34], and SVM, respectively holds less F-measure value. The MADDHM model attained the maximum NPV (approximately 0.94) with superior outcomes for detecting autism at 90 % of the learning percentage. The MCC of the MADDHM model achieved an MCC value of 0.948, a significant increase compared to the MCC values of Bi-GRU, CNN, DBN [46], DNN [47], LSTM, VGG19 [33], ResNet50 [34], and SVM, which ranged from 0.7 to 0.8. Furthermore,

for dataset 2, the MADDHM model attained a higher F-measure value (0.94) at 90 percent of the learning compared to the existing schemes. Therefore, the MADDHM model significantly holds larger outcomes for detecting autism more effectively by means of pertinent features from the images and EEG signals.

4.7. Ablation study of MADDHM vs existing schemes for detecting autism

Table 4 compares the ablation study of the MADDHM approach for detecting autism to existing techniques such as hybrid model without feature extraction, hybrid model with conventional SSE, hybrid model

Table 5

Analysis of the overall performance of the MADDHM model to existing approaches

Measures	Learning percentage =60 %								
	MADDHM	Bi-GRU	CNN	DBN	DNN	LSTM	SVM	ResNet50	VGG19
Sensitivity	0.944	0.865	0.846	0.893	0.872	0.862	0.866	0.859	0.877
Specificity	0.930	0.855	0.921	0.926	0.881	0.924	0.838	0.891	0.861
Accuracy	0.948	0.908	0.821	0.878	0.849	0.855	0.914	0.886	0.907
Precision	0.952	0.916	0.836	0.927	0.902	0.840	0.846	0.837	0.836
F-measure	0.947	0.887	0.862	0.924	0.890	0.884	0.823	0.857	0.836
MCC	0.937	0.873	0.923	0.755	0.827	0.876	0.831	0.845	0.847
NPV	0.939	0.855	0.859	0.867	0.821	0.864	0.878	0.871	0.872
FPR	0.069	0.144	0.078	0.073	0.118	0.075	0.161	0.108	0.138
FNR	0.055	0.134	0.153	0.106	0.127	0.137	0.133	0.140	0.122
Learning percentage=70									
Measures	MADDHM	Bi-GRU	CNN	DBN	DNN	LSTM	SVM	ResNet50	VGG19
Sensitivity	0.928	0.828	0.846	0.868	0.848	0.820	0.848	0.834	0.840
Specificity	0.922	0.848	0.782	0.887	0.863	0.887	0.886	0.886	0.885
Accuracy	0.936	0.839	0.894	0.844	0.825	0.818	0.800	0.808	0.804
Precision	0.924	0.792	0.844	0.887	0.881	0.805	0.793	0.798	0.795
F-measure	0.929	0.783	0.892	0.887	0.872	0.844	0.838	0.840	0.839
MCC	0.920	0.893	0.838	0.855	0.837	0.828	0.819	0.823	0.820
NPV	0.930	0.883	0.819	0.844	0.800	0.844	0.828	0.835	0.831
FPR	0.103	0.128	0.173	0.156	0.175	0.261	0.149	0.204	0.176
FNR	0.094	0.149	0.184	0.189	0.137	0.168	0.204	0.185	0.194
Learning percentage=80									
Measures	MADDHM	Bi-GRU	CNN	DBN	DNN	LSTM	SVM	ResNet50	VGG19
Sensitivity	0.930	0.837	0.837	0.867	0.850	0.833	0.874	0.854	0.864
Specificity	0.932	0.791	0.828	0.885	0.867	0.900	0.887	0.893	0.890
Accuracy	0.943	0.836	0.835	0.843	0.826	0.892	0.892	0.892	0.892
Precision	0.938	0.873	0.872	0.885	0.884	0.825	0.846	0.836	0.841
F-measure	0.932	0.818	0.748	0.885	0.875	0.861	0.817	0.839	0.828
MC	0.929	0.800	0.827	0.817	0.893	0.865	0.846	0.856	0.851
NPV	0.935	0.877	0.837	0.843	0.801	0.848	0.832	0.840	0.836
FPR	0.092	0.163	0.173	0.156	0.173	0.254	0.162	0.208	0.185
FNR	0.080	0.138	0.192	0.114	0.132	0.130	0.138	0.134	0.136
Learning percentage =90									
Measures	MADDHM	Bi-GRU	CNN	DBN	DNN	LSTM	SVM	ResNet50	VGG19
Sensitivity	0.938	0.843	0.837	0.864	0.852	0.814	0.857	0.836	0.847
Specificity	0.937	0.892	0.848	0.881	0.869	0.881	0.817	0.849	0.833
Accuracy	0.948	0.817	0.896	0.842	0.826	0.816	0.889	0.853	0.871
Precision	0.940	0.899	0.817	0.881	0.886	0.796	0.900	0.848	0.874
F-measure	0.93	0.828	0.837	0.883	0.877	0.836	0.854	0.845	0.850
MCC	0.939	0.810	0.876	0.869	0.892	0.866	0.800	0.833	0.816
NPV	0.936	0.790	0.799	0.842	0.802	0.842	0.789	0.815	0.802
FPR	0.079	0.123	0.142	0.157	0.173	0.263	0.137	0.200	0.169
FNR	0.062	0.143	0.128	0.118	0.130	0.162	0.123	0.1439	0.133

with conventional AAM feature, hybrid model with conventional SLBT feature, unimodal features for face, unimodal features for EEG, without pre-processing and hybrid model with conventional feature level fusion for dataset1 and dataset2. Moreover, the MADDHM scheme achieves the highest sensitivity (0.927) for detecting autism whilst the other schemes generate comparatively lower specificity values. The result demonstrates that the suggested model scores better result under all the metrics for detecting autism compared to other models. Therefore, the MADDHM (hybrid classifier with improved features) scheme boosts the effectiveness of detecting autism based on the images and EEG signals.

4.8. Overall performance analysis of MADDHM model vs traditional Schemes for detecting autism

For learning percentages 60, 70, 80, and 90, Tables 5 and 6 evaluates the overall performance of the MADDHM model over other models, including Bi-GRU, CNN, DBN [46], DNN [47], VGG19 [33], ResNet50 [34], LSTM, and SVM, respectively for detecting the autism regarding dataset1 and dataset2. The MADDHM model achieved minimal negative measure values (FNR, FPR) with the best outcomes compared to other existing models for detecting autism. Likewise, the MADDHM model

achieved larger MCC values(0.92) for detecting autism than existing models like Bi-GRU, CNN, DBN [46], DNN [47], LSTM, VGG19 [33], ResNet50 [34] and SVM, respectively at a learning percentage of 70 %. Moreover, the MADDHM model attains a higher accuracy value (0.948) for detecting autism at a learning percentage of 90 % than at a learning percentage of 60 % for dataset1 compared to the other models. The MADDHM scheme achieved a higher sensitivity value of 0.958 at 90 % of the learning of dataset2 whilst the existing approaches scored poor value on all measures compared to the suggested MADDHM model. This demonstrated the improvement of the MADDHM model shows outstanding performance in the autism detection process using the images and EEG signals.

4.9. Analysis-based accuracy

Table 7 shows the EEG and Face analysis based on accuracy for dataset1 and dataset2. The suggested model has examined the detection accuracy based on EEG signals and facial images. As a result, the MADHDM scheme attained 0.91 accuracy on EEG signal and 0.916 accuracy on facial image-based autism identification. For dataset 2, the MADHDM model achieved 0.935 accuracy on facial image-based autism

Table 6

Analysis of overall performance of MADDHM model to existing approaches for dataset2.

Learning percentage= 60									
Measures	MADDHM	Bi-GRU	CNN	DBN	DNN	LSTM	SVM	ResNet50	VGG19
Sensitivity	0.9446	0.8659	0.8462	0.8933	0.8723	0.8630	0.8660	0.8592	0.8775
Specificity	0.9307	0.8558	0.9219	0.9262	0.8820	0.9241	0.8389	0.8916	0.8619
Accuracy	0.9487	0.9081	0.8217	0.8788	0.8491	0.8556	0.9142	0.8864	0.9078
Precision	0.9523	0.9163	0.8370	0.9273	0.9021	0.8403	0.8465	0.8370	0.8362
F-measure	0.9478	0.8877	0.8627	0.9248	0.8904	0.8848	0.8239	0.8572	0.8368
MCC	0.9375	0.8730	0.9235	0.7559	0.8277	0.8762	0.8318	0.8457	0.8474
NPV	0.9394	0.8554	0.8596	0.8670	0.8218	0.8647	0.8788	0.8714	0.8728
FPR	0.0693	0.1442	0.0781	0.0738	0.1180	0.0759	0.1611	0.1084	0.1381
FNR	0.0554	0.1341	0.1538	0.1067	0.1277	0.1370	0.1340	0.1408	0.1225

Learning percentage=70									
Measures	MADDHM	Bi-GRU	CNN	DBN	DNN	LSTM	SVM	ResNet50	VGG19
Sensitivity	0.9544	0.8515	0.8788	0.8868	0.8733	0.8393	0.8665	0.8648	0.8618
Specificity	0.9570	0.8806	0.8082	0.9090	0.8864	0.9087	0.9069	0.9084	0.9110
Accuracy	0.9711	0.8607	0.9164	0.8608	0.8492	0.8434	0.8281	0.8364	0.8320
Precision	0.9611	0.8092	0.8758	0.9214	0.9004	0.8365	0.8182	0.8289	0.8273
F-measure	0.9550	0.8118	0.9217	0.9180	0.8958	0.8611	0.8558	0.8655	0.8592
MCC	0.9399	0.9148	0.8697	0.8861	0.8699	0.8456	0.8403	0.8549	0.8467
NPV	0.9634	0.9075	0.8375	0.8712	0.8205	0.8605	0.8514	0.8650	0.8638
FPR	0.0430	0.1194	0.1918	0.0910	0.1136	0.0913	0.0931	0.0916	0.0890
FNR	0.0456	0.1485	0.1212	0.1132	0.1267	0.1607	0.1335	0.1352	0.1382

Learning percentage=80									
Measures	MADDHM	Bi-GRU	CNN	DBN	DNN	LSTM	SVM	ResNet50	VGG19
Sensitivity	0.9555	0.8571	0.8665	0.8904	0.8708	0.8630	0.8945	0.8862	0.8841
Specificity	0.9618	0.8094	0.8597	0.9128	0.8955	0.9199	0.9058	0.9190	0.9186
Accuracy	0.9665	0.8616	0.8656	0.8664	0.8558	0.9204	0.9269	0.9174	0.9254
Precision	0.9692	0.9071	0.9030	0.9060	0.9081	0.8528	0.8701	0.8636	0.8590
F-measure	0.9639	0.8489	0.7735	0.9058	0.8996	0.8787	0.8415	0.8625	0.8612
MCC	0.9577	0.8173	0.8543	0.8452	0.9217	0.8852	0.8681	0.8852	0.8775
NPV	0.9718	0.8973	0.8588	0.8680	0.8203	0.8674	0.8525	0.8627	0.8692
FPR	0.0383	0.1906	0.1403	0.0872	0.1045	0.0801	0.0942	0.0810	0.0814
FNR	0.0383	0.1906	0.1403	0.0872	0.1045	0.0801	0.0942	0.0810	0.0814

Learning percentage=90									
Measures	MADDHM	Bi-GRU	CNN	DBN	DNN	LSTM	SVM	ResNet50	VGG19
Sensitivity	0.9584	0.8661	0.8563	0.8876	0.8698	0.8411	0.8862	0.8559	0.8750
Specificity	0.9636	0.9245	0.8768	0.9126	0.8985	0.9009	0.8387	0.8759	0.8504
Accuracy	0.9693	0.8411	0.9240	0.8716	0.8444	0.8486	0.9137	0.8797	0.8986
Precision	0.9638	0.9230	0.8422	0.9047	0.9141	0.8142	0.9264	0.8817	0.8928
F-measure	0.9646	0.8510	0.8600	0.9054	0.9050	0.8551	0.8836	0.8723	0.8740
MCC	0.9754	0.8376	0.8981	0.8887	0.9134	0.8968	0.8258	0.8588	0.8420
NPV	0.9725	0.8215	0.8167	0.8734	0.8344	0.8697	0.8072	0.8391	0.8296
FPR	0.0364	0.0755	0.1232	0.0874	0.1015	0.0991	0.1613	0.1241	0.1496
FNR	0.0416	0.1339	0.1437	0.1124	0.1302	0.1589	0.1138	0.1441	0.1250

Table 7

EEG and face analysis for dataset1 and dataset2.

Dataset 1			Dataset 2		
Methods	EEG Accuracy	Face Accuracy	Methods	EEG accuracy	Face accuracy
MADDHM	0.9104	0.9167	MADDHM	0.9286	0.9351
Bi_GRU	0.8992	0.9003	Bi_GRU	0.9172	0.9183
CNN	0.8674	0.8203	CNN	0.8847	0.8367
DBN	0.7999	0.8477	DBN	0.8159	0.8646
DNN	0.8284	0.8860	DNN	0.8450	0.9037
LSTM	0.8847	0.8129	LSTM	0.9024	0.8292
SVM	0.9014	0.8749	SVM	0.9194	0.8924
ResNet50	0.8930	0.8439	ResNet50	0.9109	0.8608
VGG19	0.8972	0.8594	VGG19	0.9152	0.8766

detection. Moreover, the developed method achieves the highest detection accuracy based on both facial images and EEG signals as compared to the other traditional schemes including Bi-GRU, CNN, DBN [46], DNN [47], LSTM, VGG19 [33], ResNet50 [34] and SVM, respectively for detecting the autism

Table 8

Friedman test.

Dataset 1		Dataset 2	
Methods	P-Value	Methods	P-value
MADDHM	0.0629	MADDHM	0.0641
Bi_GRU	0.0538	BiGRU	0.0549
CNN	0.0585	CNN	0.0596
DBN	0.0433	DBN	0.0441
DNN	0.0498	DNN	0.0508
LSTM	0.0328	LSTM	0.0334
SVM	0.0385	SVM	0.0392
ResNet50	0.0356	ResNet50	0.0363
VGG19	0.0370	VGG19	0.0378

4.10. Statistical test

Tables 8 and 9 display the Friedman and Wilcoxon test of p-value. Nonparametric statistical techniques, often known as distribution-free techniques, do not rely on the presumption that the data are taken from a specific probability distribution. The Wilcoxon signed-rank test is

Table 9
Wilcoxon test.

Dataset 1		Dataset 2	
Methods	P-value	Methods	P-value
MADDHM	0.0346	MADDHM	0.0351
Bi GRU	0.0599	Bi GRU	0.0608
CNN	0.0768	CNN	0.0780
DBN[0.0468	DBN	0.0475
DNN	0.0414	DNN	0.0420
LSTM	0.0395	LSTM	0.0401
SVM	0.0684	SVM	0.0694
ResNet50	0.0539	ResNet50	0.0547
VGG19	0.0612	VGG19	0.0621

Table 10
Analysis of the P-test and T-test of the MADDHM over the existing models.

Dataset 1			Dataset 2		
Methods	P-test	T-test	Methods	P test	T-test
MADDHM	0.3849	0.3018	MADDHM	0.3906	0.3064
Bi GRU	0.8949	0.4859	Bi GRU	0.9083	0.4932
CNN	0.6480	0.6884	CNN	0.6577	0.6987
DBN	0.7003	0.8049	DBN	0.7108	0.8169
DNN	0.8294	0.7295	DNN	0.8418	0.7404
LSTM	0.4549	0.5839	LSTM	0.4618	0.5927
SVM	0.5750	0.3948	SVM	0.5836	0.4007
ResNet50	0.6198	0.5694	ResNet50	0.6291	0.5779
VGG19	0.5499	0.5160	VGG19	0.5582	0.5238

used for paired samples, whereas the Friedman test is used for independent samples. The Wilcoxon rank-sum test is used to compare two independent samples.

4.11. P-test and T-test analysis

The t-statistic quantifies the distinction between two sets by representing it in terms of standard error units. A low P-value also signifies significance as it indicates the probability of an observation falling within extreme t-values. Table 10 compares the analysis of the P-test and T-test of the MADDHM over the traditional models such as Bi-GRU, LSTM, SVM, CNN, DNN, ResNet50 [33], VGG19 [34], and DBN for dataset1 and dataset 2. The MADDHM offers a better value on the P-test about 0.384 for dataset1 and the T-test about 0.306 for dataset2 compared to the other existing models. Therefore, the suggested MADDHM model achieved better performance results on autism detection compared to the traditional models.

4.12. K-Fold validation analysis

In applied machine learning, cross-validation is mostly used to gauge a machine learning model's proficiency on hypothetical data. Fig 13 shows the analysis of K-fold validation. That is, to assess the model's projected overall performance with a small sample size when it comes to making predictions on data that wasn't used for model training. It is a widely used strategy because, in comparison to other approaches, such as a straightforward train/test split, it typically yields a less biased or optimistic estimate of the model skill and is easy to grasp.

5. Discussion

Our proposed approach could enhance recognition ability by learning complementary information between two different modalities. Additionally, it makes it possible to jointly model and analyze facial and EEG data using a variety of analytic techniques, including accuracy-based analysis, statistical test analysis, ablation analysis, and analysis using the P- and T-tests. Experimental results with an overall accuracy of 89.26 % demonstrated that our proposed strategy surpassed multimodal approaches and a basic feature-level fusion method in terms of identification performance. Moreover, in terms of precision outcomes for autism identification, the MADDHM model beats other techniques including Bi-GRU, CNN, DBN [46], DNN [47], LSTM, SVM, ResNet50 [33], and VGG19 [34] at a learning percentage of 70 %. In comparison to the existing techniques, the MADDHM model achieved a higher rate of precision at 90 % of learning about 0.940 meanwhile Bi-GRU (0.899), CNN (0.817), DBN (0.881), DNN (0.886), LSTM (0.796), SVM (0.900), ResNet50 [33] (0.848), and VGG19 [34] (0.874). As a result, when it came to autism detection, the recommended MADDHM model outperformed the conventional models.

Additionally, the detection accuracy of the proposed model is assessed using facial photographs and EEG signals. As a result, the MADDHM method achieved 0.916 accuracy for face image-based autism identification and 0.91 accuracy for EEG signals, while existing models including CNN, Bi-GRU, DBN [46], DNN [47], LSTM, ResNet50 [33], and VGG19 [34] achieved lower accuracy values. Hence, as compared to other models, the MADDHM model performs exceptionally well in multimodal autism identification. This new model enables clinicians to diagnose children with ASD more objectively and reliably, improving diagnosis efficiency and demonstrating the instrument's immense potential for clinical applications. This suggested distribution will also be very beneficial in recognizing the short-duration EEG sample signals and in understanding human emotions.

6. Conclusion

This article has proposed a MADDHM model considering two modalities Face and EEG. The proposed autism detection framework

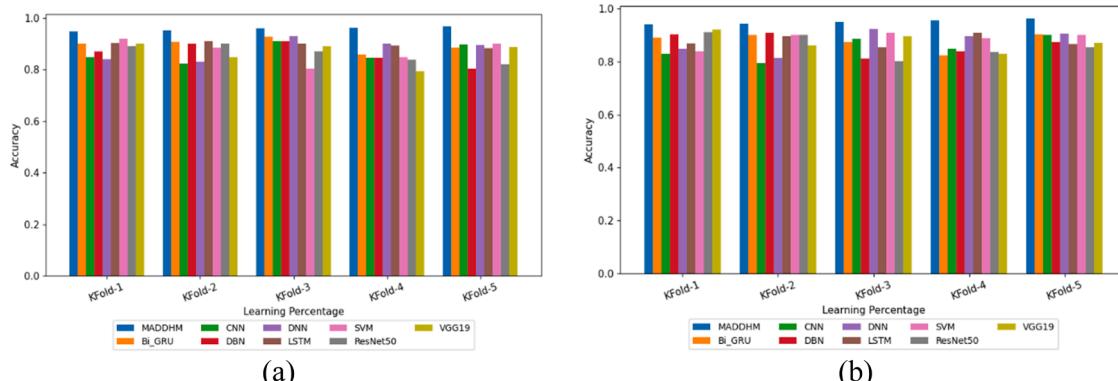


Fig 13. Analysis of K-Fold validation for datasets 1 and 2.

includes the following phases: Pre-processing, Feature Extraction, Improved Feature level Fusion, and Detection. During the pre-processing stage, the input EEG signal was pre-processed by Wiener filtering and the face image was pre-processed using Gabor filtering. From the EEG signal data, features like Improved Singular Spectrum Entropy-based features, correlation dimension, and CSP-based features were extracted. From the face image, features will be extracted like Improved Active Appearance model-based features, Proposed SLBT, and GLCM features as well. After the feature extraction, they were fused to form the combined feature set by following the Improved Feature level fusion process. Based on the fused features, the detection process takes place by the hybrid model that combines models like CNN and Bi-GRU. Thus, the accepted scheme was evaluated with alternative approaches by a variety of measures. The accuracy of the proposed MADDHM model was 91.03 % for EEG and 91.67 % for face analysis, while SVM, DNN, Bi-GRU, LSTM, and CNN were 87.02 %, 87.59 %, and 87.02 %, respectively. Moreover, the suggested MADDHM model attained higher values on specificity and MCC of about 0.937 and 0.939 respectively which surpasses the result of the traditional techniques. Therefore, the suggested model exhibits outstanding performance in detecting autism.

On the other hand, the simultaneous availability of multimodal data from both ET and EEG modalities is necessary for the practical implementation of our approach. However, in the proposed approach, extraction of a diverse set of features from EEG signals and facial images can be computationally intensive and combining these numerous features from both modalities can lead to a high-dimensional feature space, which may increase the risk of overfitting. Additionally, the development of reliable and accurate models becomes difficult in the absence of adequate multimodal data. The availability of large, diverse datasets is critical for training accurate models. In order to overcome these constraints, we will look into more effective models in the future by utilizing cutting-edge neural network techniques, like convolution neural networks (CNN) and attention networks, to fuse multimodal data especially when one modality is absent during the model training process. Additionally, we'll try to use cutting-edge signal processing techniques to investigate more useful ASD traits. Additionally, there is an expensive computational cost when dealing with real-time data. We will gather real-time autism data from various hospitals and extent our in the future and investigate the performance of the model.

Funding Information

None.

CRediT authorship contribution statement

S. Vidivelli: Writing – review & editing, Writing – original draft. **P. Padmakumari:** Visualization. **P. Shanthi:** Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] M.S. Farooq, R. Tehseen, M. Sabir, Z. Atal, Detection of autism spectrum disorder (ASD) in children and adults using machine learning, *Sci. Rep.* 13 (1) (2023) 9605.
- [2] S. Guruvammal, T. Chellatamilan, L.J. Deborah, Optimal feature selection and hybrid classification for autism detection in young children, *Comput. J.* 64 (11) (2019) 1760–1774, <https://doi.org/10.1093/comjnl/bxaa156>.
- [3] O. Peña, F.L. Cibrian, M. Tentori, Circus in Motion: a multimodal exergame supporting vestibular therapy for children with autism, *J. Multimodal User Interf.* 15 (2021) 283–299, <https://doi.org/10.1007/s12193-020-00345-9>.
- [4] J. Zhang, Y. Min, Four-classes human emotion recognition via entropy characteristic and random Forest, *Inf. Technol. Control.* 49 (3) (2020) 285–298.
- [5] V. Yaneva, L.A. Ha, S. Eraslan, Y. Yesilada, R. Mitkov, Detecting high-functioning autism in adults using eye tracking and machine learning, *IEEE Trans. Neural Syst. Rehabil. Eng.* 28 (6) (2020) 1254–1261, <https://doi.org/10.1109/TNSRE.2020.2991675>.
- [6] S. Sarabadiani, L.C. Schudlo, A.A. Samadani, A. Kushski, Physiological detection of affective states in children with autism spectrum disorder, *IEEE Trans. Affective Comput.* 11 (4) (2020) 588–600, <https://doi.org/10.1109/TAFFC.2018.2820049>.
- [7] X. Yang, M.-L. Shyu, H.-Q. Yu, S.-M. Sun, N.-S. Yin, W. Chen, Integrating image and textual information in human–robot interactions for children with autism spectrum disorder, *IEEE Trans. Multimedia* 21 (3) (2019) 746–759, <https://doi.org/10.1109/TMM.2018.2865828>.
- [8] G. Tan, K. Xu, J. Liu, H. Liu, A Trend on autism spectrum disorder research: eye tracking–EEG correlative analytics, *IEEE Transact. Cognit. Developm. Syst.* 14 (3) (2022) 1232–1244, <https://doi.org/10.1109/TCDS.2021.3102646>. Sept.
- [9] M.T. Tomczak, et al., Stress monitoring system for individuals with autism spectrum disorders, *IEEE Access* 8 (2020) 228236–228244, <https://doi.org/10.1109/ACCESS.2020.3045633>.
- [10] N. Rusli, S.N. Sidek, H.M. Yusof, N.I. Ishak, M. Khalid, A.A.A. Dzulkarnain, Implementation of wavelet analysis on thermal images for affective states recognition of children with autism spectrum disorder, *IEEE Access* 8 (2020) 120818–120834, <https://doi.org/10.1109/ACCESS.2020.3006004>.
- [11] T. Akter, et al., Machine learning-based models for early stage detection of autism spectrum disorders, *IEEE Access* 7 (2019) 166509–166527, <https://doi.org/10.1109/ACCESS.2019.2952609>.
- [12] C.L. Alves, T.G.D.O. Toutain, P. de Carvalho Aguiar, A.M. Pineda, K. Roster, C. Thielemann, J.A.M. Porto, F.A. Rodrigues, Diagnosis of autism spectrum disorder based on functional brain networks and machine learning, *Sci. Rep.* 13 (1) (2023) 8072.
- [13] S.K. Jarraya, M. Masmoudi, M. Hammami, Compound emotion recognition of autistic children during meltdown crisis based on deep spatio-temporal analysis of facial geometric features, *IEEE Access* 8 (2020) 69311–69326, <https://doi.org/10.1109/ACCESS.2020.2986654>.
- [14] Pramit MazumdarGiuliano ArruFederica Battisti, Early detection of children with Autism Spectrum Disorder based on visual exploration of images, *Signal Process. Image Commun.* 94 (2021), 10 February(Cover date: May 2021)Article 116184.
- [15] M.J. Uddin, M.M. Ahamed, P.K. Sarker, S. Aktar, N. Alotaibi, S.A. Alyami, M. A. Kabir, M.A. Moni, An integrated statistical and clinically applicable machine learning framework for the detection of autism spectrum disorder, *Computers* 12 (5) (2023) 92.
- [16] Judith CharpentierMarianne LatinusMarie Gomot, Brain correlates of emotional prosodic change detection in autism spectrum disorder, *NeuroImage: Clinical* 28 (2020), 27 November(Cover date: 2020)Article 102512.
- [17] Nikita GoelBhavya GroverMoolchand Sharma, Modified grasshopper optimization algorithm for detection of autism spectrum disorder, *Phys. Commun.* 41 (2020) (Cover date: August 2020)Article 101115.
- [18] Tong WuHongchao WangWei Chen, Potential of gut microbiome for detection of autism spectrum disorder, *Microb. Pathog.* 149 (2020), 20 October(Cover date: December 2020)Article 104568.
- [19] U. Narasimhan, R. Rajendran, D.A. Abraham, et al., Prevalence and pattern of complementary and alternative medicine for autism spectrum disorder in tamil nadu, *Indian J. Pediatr.* 87 (2020) 400, <https://doi.org/10.1007/s12098-019-03142-9>.
- [20] Jayne Jennings Dunlap, Autism spectrum disorder screening and early action, *J. Nurse Practition.* 15 (7) (2019) 496–501 (Cover date: July–August 2019).
- [21] B. Ku, J.D. Stinson, M. MacDonald, Parental behavior comparisons between parents of children with autism spectrum disorder and parents of children without autism spectrum disorder: a meta-analysis, *J. Child Fam. Stud.* 28 (2019) 1445–1460, <https://doi.org/10.1007/s10826-019-01412-w>.
- [22] ... & N.M. Krishna, K. Sekaran, A.V.N. Vamsi, G.P. Ghantasala, P. Chandana, S. Kadry, R. Damasevičius, An efficient mixture model approach in brain-machine interface systems for extracting the psychological status of mentally impaired persons using EEG signals, *Ieee Access* 7 (2019) 77905–77914.
- [23] T.M. Ghazal, S. Munir, S. Abbas, A. Athar, H. Alrababah, M.A. Khan, Early detection of autism in children using transfer learning, *Intellig. Automat. Soft Comput.* 36 (1) (2023) 11–22.
- [24] J. Han, G. Jiang, G. Ouyang, X. Li, A multimodal approach for identifying autism spectrum disorders in children, *IEEE Trans. Neural Syst. Rehabil. Eng.* 30 (2022) 2003–2011.
- [25] M. Eni, I. Dinstein, M. Ilan, I. Menashe, G. Meiri, Y. Zigel, Estimating autism severity in young children from speech signals using a deep neural network, *IEEE Access* 8 (2020) 139489–139500, <https://doi.org/10.1109/ACCESS.2020.3012532>.
- [26] K. Barik, K. Watanabe, J. Bhattacharya, et al., A fusion-based machine learning approach for autism detection in young children using magnetoencephalography signals, *J. Autism Dev. Disord.* (2022), <https://doi.org/10.1007/s10803-022-05767-w>.
- [27] K. Vakadkar, D. Purkayastha, D Krishnan, Detection of autism spectrum disorder in children using machine learning techniques, *Sn Comput.. Sci.* 2 (2021) 386, <https://doi.org/10.1007/s42979-021-00776-5>.
- [28] A.A. Ramírez-Duque, A. Frizera-Neto, T.F. Bastos, Robot-assisted autism spectrum disorder diagnostic based on artificial reasoning, *J. Intell. Robot Syst.* 96 (2019) 267–281, <https://doi.org/10.1007/s10846-018-00975-y>.
- [29] A. Saranya, R. Anandan, FIGS-DEAF: an novel implementation of hybrid deep learning algorithm to predict autism spectrum disorders using facial fused gait features, *Distrib Parallel Databases* 40 (2022) 753–778, <https://doi.org/10.1007/s10619-021-07361-y>.

- [30] P. Lu, X. Li, L. Hu, et al., Integrating genomic and resting State fMRI for efficient autism spectrum disorder classification, *Multimed Tools Appl.* 81 (2022) 19183–19194, <https://doi.org/10.1007/s11042-020-10473-9>.
- [31] Puli, A. Kushki, Toward Automatic anxiety detection in autism: a real-time algorithm for detecting physiological arousal in the presence of motion, *IEEE Trans. Biomed. Eng.* 67 (3) (2020) 646–657, <https://doi.org/10.1109/TBME.2019.2919273>.
- [32] K. Devika, D. Mahapatra, R. Subramanian, V.R.M. Oruganti, Outlier-based autism detection using longitudinal structural MRI, *IEEE Access* 10 (2022) 27794–27808, <https://doi.org/10.1109/ACCESS.2022.3157613>.
- [33] F.W. Alsaade, M.S. Alzahrani, Classification and detection of autism spectrum disorder based on deep learning algorithms, *Comput. Intell. Neurosci.* 2022 (2022).
- [34] S. Akter, H. Shahriar, A. Cuzzocrea, Autism disease detection using transfer learning techniques: performance comparison between central processing unit vs graphics processing unit functions for neural networks, in: 2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC), IEEE, 2023, pp. 1084–1092.
- [35] A.K. Kareem, M.M. AL-Ani, A.A. Nafea, Detection of autism spectrum disorder using a 1-dimensional convolutional neural network, *Baghdad Sci. J.* 20 (3 (Suppl.)) (2023), 1182–1182.
- [36] B. Awaji, E.M. Senan, F. Olayah, E.A. Alshari, M. Alsulami, H.A. Abosaq, J. Alqahtani, P. Janrao, Hybrid techniques of facial feature image analysis for early detection of autism spectrum disorder based on combined CNN features, *Diagnostics* 13 (18) (2023) 2948.
- [37] Masoud Hassan, Haval Hussein, Adel Eesa, Ramadhan Mstafa, Face Recognition Based on Gabor Feature Extraction Followed by FastICA and LDA, *Comput. Mater. Continua* 68 (2021) 1637–1659, <https://doi.org/10.32604/cmc.2021.016467>.
- [38] Xiaoan & Yan, Ying & Liu, Yadong Xu, Minping Jia, Multistep forecasting for diurnal wind speed based on hybrid deep learning model with improved singular spectrum decomposition, *Energy Convers. Manage.* 225 (2020) 113456, <https://doi.org/10.1016/j.enconman.2020.113456>.
- [39] Harikumar Rajaguru, V. Arunachalam, Correlation dimension based performance analysis of alcoholic EEG data with PCA and PSO classifiers, *Int. J. Recent Technol. Eng.* 7 (2018) 403–406.
- [40] X. Jiang, X. Gu, Z. Mei, H. Ren, W. Chen, A modified common spatial pattern algorithm customized for feature dimensionality reduction in fNIRS-Based BCIs, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 2018, pp. 5073–5076, <https://doi.org/10.1109/EMBC.2018.8513454>.
- [41] Tao Zhou, Xiao-Jun Wu, Tao Wu, Zhen-Hua Feng, An improved AAM method for extracting human facial features, *J. Appl. Math.* (2012) 10, <https://doi.org/10.1155/2012/643562>. VolumeArticle ID 643562pages.
- [42] John Ojo, Aborisade, Abraham Amole, Adewumi Durodola, Comparative analysis of textural features derived from GLCM for ultrasound liver image classification, *Int. J. Comput. Trends and Technol.* 11 (2014), <https://doi.org/10.14445/22312803/IJCTT-V11P151>.
- [43] N.S. LakshmiPrabha, S. Majumder, Face recognition system invariant to plastic surgery, in: 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA), Kochi, India, 2012, pp. 258–263, <https://doi.org/10.1109/ISDA.2012.6416547>.
- [44] Di ZhaoJian WangYijia Zhang, Extracting drug–drug interactions with hybrid bidirectional gated recurrent unit and graph convolutional network, *J. Biomed. Inform.* 99 (2019), 27 September(Cover date: November 2019) Article 10329.
- [45] Y. LeCun, K. Kavukcuoglu, C. Farabet, Convolutional networks and applications in vision, in: *Circuits and Systems, International Symposium on*, 2010, pp. 253–256.
- [46] Z.-A. Huang, Z. Zhu, C.H. Yau, K.C. Tan, Identifying autism spectrum disorder from resting-state fMRI using deep belief network, *IEEE Transact. Neural Networks and Learn. Syst.* 32 (7) (2021) 2847–2861, <https://doi.org/10.1109/TNNLS.2020.3007943>. July.
- [47] K K. Mujeeb Rahman, M. Subashini, Identification of autism in children using static facial features and deep neural networks, *Brain Sci.* 12 (2022) 94, <https://doi.org/10.3390/brainsci12010094>.