JSCDM

JOURNAL OF SOFT COMPUTING AND DATA MINING
e-ISSN: 2716-621X

Vol. 7 No. 2 (2026) 371-387
https://publisher.uthm.edu.my/ojs/index.php/jscdm

A Multi-Model Ensemble Approach Using Deep and Traditional Learning for Autism Spectrum Disorder Classification

Dhafar Fakhry Hasan[1]*, Maha A. Abdul-Jabar[2], Mawadah Mohammed Suliman[3], Massila Kamalrudin[4]

[1] *Supporting Science Unit, College of Medicine, University of Mosul, Mosul 41012, Nineveh, IRAQ*
[2] *Computer Unit, Al-Batoul College of Medicine, University of Mosul, Mosul 41012, Nineveh, IRAQ*
[3] *College of Science, University of Mosul, Mosul 41012, Nineveh, IRAQ*
[4] *Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, 76100, Melaka, MALAYSIA*

*Corresponding Author: dhf@uomosul.edu.iq
DOI: https://doi.org/10.30880/jscdm.2026.07.02.020

Article Info

Received: 10 January 2026
Accepted: 29 May 2026
Available online: 30 June 2026

Keywords

Autism Spectrum Disorder (ASD), classification, ensemble learning, deep learning, machine learning, weighted majority voting.

Abstract

Conventional diagnostic tests for Autism Spectrum Disorder (ASD) involve the use of subjective behavioral observations and questionnaires completed by the clinician, which can be time-consuming and subject to human bias. The challenge encourages the development of innovative, data-driven methods to facilitate early and accurate identification of ASD. The research proposes a Multi-Model Ensemble Approach Using Deep and traditional learning for ASD classification (MME-ASD) model. The MME-ASD model encompasses three traditional machine learning (ML) and two deep learning (DL) algorithms that perform according to a weighted majority voting strategy. The five learning paradigms are Random Forest (RF), Decision Tree (DT), Neural Networks (NN), Convolutional Neural Networks (CNN), and Deep Recurrent Neural Networks (DRNN), which are utilized to enhance classification accuracy and generalization. An ensemble evaluation method is proposed to complete this study and assess the efficiency of the proposed MME-ASD model. The MME-ASD model acquires complementary properties by using numeric and textual data from a publicly available dataset of ASD, which includes information on 704 adults, both with and without a diagnosis. Initially, during the evaluation phase, the performance of the standalone traditional ML and DL algorithms was assessed across several train-test ratios. Subsequently, the proposed MME-ASD ensemble was evaluated with a 60-40 split to ensure compatibility with the baseline models. Finally, a 3-fold cross-validation experiment was conducted to assess the robustness and generalization of the proposed MME-ASD model. The experimental outcomes reveal that the MME-ASD model outperforms individual learners for both cross-validation and train-test assessments. It records evaluation metrics of accuracy 99.57%, precision 99.48%, and recall 98.94% across the 3-fold cross-validation experiments. The findings verify that incorporating deep and traditional learning models in an ensemble framework can significantly enhance the classification of ASD, offering a dependable and scalable computational model to aid clinical specialists in the initial diagnosis of ASD.

1. Introduction

Autism spectrum disorder, often known as ASD, is a neurodevelopmental condition in which a person's relationships and communication with other people are impaired throughout the remainder of their life. Autism is considered a "behavioral disorder" even though it may be diagnosed at any age [1]. Autism symptoms often appear during the first two years of a person's life. ASD makes it difficult for a person to utilize their linguistic, communicative, cognitive, and social abilities. Recent research, such as [2], in the field of behavioral sciences has explored the potential of utilizing intelligent computer technologies based on Machine Learning (ML) to reduce the time required for testing or enhance diagnostic sensitivity, specificity, and accuracy.

In recent years, ASD has been explored by social and computational information scientists, making use of cutting-edge technologies such as Deep Learning (DL) and ML [3]. Their goal is to improve the diagnostic timeliness, accuracy, and efficiency of the test. The field of study known as ML employs ingenious methods to uncover hidden, valuable patterns, which may then be used to improve decision-making processes [4]. The performance of instruments may be improved with the assistance of ML, enhancing their adaptability to new data and reducing the algorithm's need for fewer lines of code. Moreover, the identification of objective biomarkers that may aid in the diagnosis and treatment of brain-based diseases such as ASD is the major purpose of research conducted in the field of psychiatric neuroimaging [5]. The use of data-intensive ML techniques to examine the replicability of brain activity patterns across newer and more precise datasets is an exciting new development.

The degree of social, communication, and sensorimotor deficiencies may vary widely among people who have ASD, which is linked with a spectrum of phenotypes [6]. Diagnostic tools for ASD look at a person's typical social behaviors as well as their linguistic abilities. However, research in the field of neuroscience may assist in bridging the gap between an accurate mapping of the complexity of the spectrum of behavioral changes associated with autism and their brain patterns [7].

Studies using non-invasive imaging techniques of the brain have led to advances in our understanding of the neural bases of brain-based disorders and the behaviors associated with them. One such disorder is ASD, which is characterized by deficits in social interaction and communication [8]. Rapid advancements in DL have enabled the integration of various types of data, including data with multiple modalities [9]. DL has been shown to be useful in the treatment of medical issues in a number of recent studies [10]. For instance, a fusion of latent feature representations that were generated from MRI and PET data has been used as a diagnostic tool for Alzheimer's disease [11], [12]. Learning difficult patterns, such as functional connectivity, is a strong suit of DL, suggesting that the technique might be useful for diagnostic purposes [13].

ASD is defined by pervasive difficulties in social communication and interaction, in addition to confined and repetitive patterns of behavior, interests, or activities. It is currently unclear what causes ASD, but some researchers have a hypothesis that the structure of the brain carries information that is important to the disorder [14]. The data include volumetric measurements and the structure of the cerebellar vermis [15], regional thicknesses taken using surface-based morphometry [16], the volumes of gray and white matter maps [17], volumetric and geometric properties extracted from chosen cortical sites, and morphometric features of selected areas of interest [18]. A few studies have revealed accuracy that is relatively high, ranging from 76% to 90%. These investigations, on the other hand, included performance evaluations of classifiers on small datasets, typically comprising fewer than 50 people [19]. Furthermore, the existing corpus of research has not yet yielded the development of reliable algorithms or demonstrated out-of-sample performance.

In recent years, there have been studies that have investigated the screening of ASD based on ML and clinical diagnostic criteria. Thabtah et al. [20] introduced a screening model for ASD based on ML adaptation and DSM-5. Their study focused on the classification of ASD using ML, and the pros and cons were discussed. The authors also highlighted a common issue that has plagued past screening tools, namely that there may be inconsistencies with screening tools that use the DSM-IV criteria instead of the newer DSM-5 guidelines. A different study [21] used classification methods for the detection of ASD. The main aim was to identify ASD and distinguish its severity levels. This approach proposed a Neural Network (NN) to combine fuzzy techniques and Support Vector Machine (SVM) for studying students' behaviors and social interactions, through the WEKA tool.

The authors of [22] presented an optimal behavioral feature set approach for detecting autism in a separate work. It is a study that features 21 features from an ASD diagnostic data set from the University of California, Irvine (UCI) Machine Learning Repository. A binary firefly feature selection wrapper was used to select the most informative features from the swarm intelligence. The hypothesis was tested to determine whether the classification accuracy of an ML model would still hold true for a small subset of features. The results indicated that 10 of the 21 features could be used to differentiate ASD cases from non-ASD cases, while 11 were not very useful. The proposed method was tested on the proposed dataset, achieving an accuracy of 92.12%-97.95%, comparable to that obtained with the full dataset.

A minimal set of behavioral features to aid in the diagnosis of autism was explored by Kosmicki et al. [23]. The authors used machine learning methods in the analysis of clinical evaluations of ASD in their study. Selected behavioral indicators were explored in children with autism-like characteristics using the Autism Diagnostic Observation Schedule (ADOS), which is comprised of four modules of the instrument. Eight different ML methods were tested, and the score-sheet data of 4,540 people were collected. The results revealed that 9 out of 28 behaviors in Module 2 and 12 out of 28 behaviors in Module 3 were sufficient to predict the risk of ASD with 98.27% and 97.66% accuracy, respectively.

Luongo et al. [24] used imitation learning with the help of ML classifiers for the identification of individuals with autism. The research objective was to investigate the issues related to discriminative testing conditions and kinematic characteristics. The sample consisted of 16 people with Autism Spectrum Condition (ASC) who participated in various hand movement activities. The researchers examined 40 kinematic features for eight imitation conditions using ML techniques. The results showed that ML methods can be valuable tools for processing high-dimensional movement data and for autism diagnosis with a small number of samples. The other thing is that the RIPPER classifier obtained different sensitivity rates with the AQ-Adolescent dataset: 87.30% with Va, 80.95% with CHI, 80.95% with IG, 84.13% with correlation-based selection, 84.13% with CFS, and 80.00% without feature selection.

The use of ML, DL, and ensemble models based on various sources of data, including EEG, MRI, facial images, health records, and behavioral screening data, has become a growing trend in recent studies on the classification of ASD between 2023 and 2025. There are still some unmet needs, however. First, most studies have been on a specific model type and/or a single data modality, limiting generalization. Second, many ensemble learning studies aggregate classifiers that are similar in nature but do not explicitly state what each learner does. Third, a few studies have focused on incorporating traditional ML and DL learners in a single weighted voting pipeline for structured ASD screening data, using the validation method. This study addresses this gap by proposing MME-ASD, which combines the RF, DT, NN, CNN, and DRNN by weighted majority voting based on the validation process and testing the model by train-test split and 3-fold cross-validation.

Although ensemble learning and weighted voting are already proven techniques, this is not an assertion of a new voting theory. Instead, the novelty is that the specific embedding of the various traditional ML and DL learners: RF, DT, NN, CNN, and DRNN is combined into a single ASD screening model by applying the validation-based weighted majority voting. This combination aims to leverage complementary learning behaviors in structured ASD screening data, where tree-based models capture rule-based feature interactions and neural models capture nonlinear, complex patterns. Hence, the primary contribution of this work is the pragmatic ensemble design, comparative assessment, and validation of this heterogeneous classification of the ASD model.

This paper primarily focuses on determining the most appropriate form of AI to use when dealing with ASD. Five traditional and deep ML algorithms were chosen for testing on the Autistic Screening Adult Dataset: RF, DT, NN, DRNN, and CNN. Moreover, the paper proposed a new Multi-Model Ensemble Approach for ASD classification (MME-ASD), which combines the five traditional and DL algorithms as a suggested ensemble architecture via weighted majority voting. Three diverse experiments were conducted on the ASD dataset: the first uses traditional standalone and deep ML algorithms with different train-split ratios. The second evaluates the proposed MME-ASD ensemble with a 60-40 split to ensure compatibility with the baseline models. Finally, a 3-fold cross-validation experiment was applied to reflect the robustness and generalization ability of the proposed ensemble. All conducted experiments are evaluated using a well-known evaluation system of measurement, namely accuracy, recall, precision, and F1-score, with an additional error rate metric used in the first experiment to reflect the error rate between the baseline traditional and DL algorithms and the proposed model.

MME-ASD is a novel methodology in three ways. First, it has a heterogeneous ensemble structure, which uses shallow ensemble tree-based learners, a standard NN, and deep neural architectures in one pipeline for ASD screening. Second, the model adopts the approach of validation-based weighting, that is, the amount of contribution that a learner makes is based on his or her validation performance, not the equal weight of the voter. Third, the framework applies both standalone learners and the final ensemble across different train-test splits and 3-fold cross-validation to assess the stability, robustness, and generalization of the proposed model. So the novelty does not reside just in the voting mechanism but in the all-in-one weighting and validation process specifically developed for the classification of ASD.

2. Methodology

In developing the MME-ASD model, we employ an ensemble evaluation methodology that organizes the work into several phases: data preparation, data preprocessing, data splitting, ML and DL modeling, initial evaluation, ensemble, prediction, and final evaluation, as shown in Figure 1. These phases ensure that all processes are carried out correctly. If something goes wrong during one of these phases, that phase can be repeated without affecting the current phases, ensuring the SAD prediction requirement is met.

The development methodology of MME-ASD begins with a thorough data preparation and data preprocessing phase to guarantee that the input data set is clean, consistent, and appropriate for multi-model learning. Data Preparation comprises the collection of features related to ASD, formatting the data, and dividing the input variables from the class label of ASD. The missing values were imputed using the mean for numerical values and the mode for categorical values. Label Encoding was used to encode binary categorical attributes (gender, jaundice, autism history, previous app usage) while One-Hot Encoding was applied to multi-class categorical attributes (ethnicity, country of residence, relation). Min-Max scaling was used to normalize the numerical attributes, such as age and screening score, between 0 and 1. The class distribution was analyzed beforehand and was found adequate, and no oversampling or under-sampling technique was employed. Imputation values, encoding mappings, and scaling parameters were fitted from the training data or training fold, and then applied to testing or validation data during train-test and cross-validation experiments, respectively. This pre-processing pipeline ensured uniform input quality across all models in the ensemble and enhanced prediction reliability and generalizability.

The second step is data splitting and modeling, which consists of splitting the processed dataset into separate training, validation and testing sets- usually based on 30:70 to 70:30. Individual ML and DL models, namely Random Forest (RF), Decision Tree (DT), Neural Networks (NN), Convolutional Neural Networks (CNN), and Deep Recurrent Neural Networks (DRNN), are developed with the use of the training subset. Both models are trained on separate data, focusing on unique representations. The ML models are trained on the relationship between tabular features, whereas the DL models learn more complex nonlinear and sequential patterns. Following training, the validation set is initially evaluated using performance metrics such as accuracy, F1-score, recall, and AUC to assess the performance of different model types.

The last step will be to compile, forecast, and conclude the analysis. In the ensemble process, the results of the five base models are fused using weighted majority voting, with each model contributing according to its validation performance. This incorporation makes it stronger because the strengths of the other mitigate the weaknesses of one model. The ensemble then makes predictions on unknown test data, generating the final ASD classification outcomes. The last evaluation step is a quantitative comparison of the ensemble's performance with that of individual models, and the conclusion is that the MME-ASD framework has better predictive capability, reliability, and stability in detecting ASD.

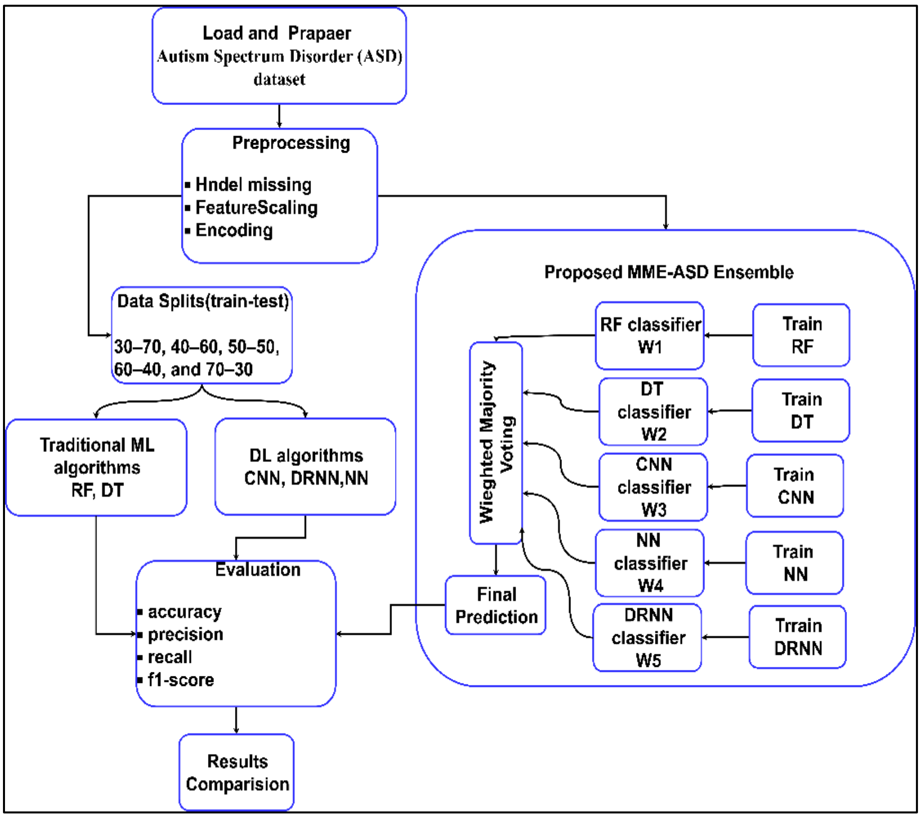


Fig. 1 *The overall process of the ensemble evaluation methodology*

2.1 Dataset

The Autism Screening Adult (ASA) dataset used in this study was from the University of California, Irvine (UCI) ML Repository. It is a classification and prediction dataset produced by Fadi Fayez Thabtah [26] as part of a project taking place at Manukau Institute of Technology. The dataset was submitted to UCI on 24 December 2017 and had a high degree of online access via the UCI platform. There are 704 adult records and 21 attributes in the ASA dataset. These attributes encompass demographic, behavioral, and screening-related. Specifically, the data includes ten items from an Autism-Spectrum Quotient (AQ) screening, age, gender, ethnicity, history of jaundice, autism in family, country of residence, use of screening application previously, screening score, age description, familial relationship, and the final ASD class label. Hence, in this study, 704 samples of 21 attributes are used. To ensure consistency, the previously mentioned value of 1100 instances was removed [26].

ASA is a structured medical and behavioral dataset intended to facilitate the diagnosis of autism in adults using demographic, medical, and behavioral signs. It has 11 attributes, which are a mixture of categorical and numerical features, describing the background of each individual and their results during the test. The patient sample is between 17 and 64 years old, which is a broad range of adults that can be applied to ASD screening. It also considers gender (Male or Female) and ethnicity, encompassing the vast number of various racial groups into White-European, South Asian, Hispanic, Asian, Middle Eastern, and others. This variance has been useful in ensuring that predictive models built on this dataset can be applied across diverse demographic profiles.

Clinical and behavioral indicators linked to ASD are also incorporated in the dataset. The Jaundice attribute captures the presence or absence of a yellowing of the skin or the eyes- an early-life medical disorder that is occasionally investigated as a developmental relationship. The Autism attribute indicates whether the individual has been diagnosed with ASD in the past, which is a powerful reference attribute for classifying their health condition. Also, the Usedappbefore field will include information about the patient who previously used the ASD screening application, indicating that she has been exposed to diagnostic instruments earlier. The result attribute is a numerical value ranging from 0 to 10 and is usually based on self-assessment responses that measure behavioral and cognitive characteristics of autism.

Other characteristics are Relation, specifying who took the test (self, a medical worker, a relative), which provides context for the reliability of the answers. The Agedesc field is a nominal indicator that the respondents are aged 18 years or older, which aligns with adult screening standards. The last characteristic is the target variable, ClassASD, which indicates whether the person is diagnosed with ASD (Yes/No) based on measures of analytical and functional connectivity. All in all, this data is an excellent mix of personal, social, and behavioral variables that can be used to train and test ML and DL models to identify ASD with high accuracy.

2.2 Algorithms

The combination of five algorithms in the MME-ASD model, including DT, RF, NN, CNN, and DRNN, is based on the idea of uniting the complementary learning properties to reach efficient and accurate ASD detection. All of the algorithms have their own analytical advantages, which, when combined via ensemble, improve the overall performance of the system and generalization. Conventional ML models such as DT and RF are highly interpretable and can be used to analyze structured tabular data, identify nonlinear relationships and interactions between features without a lot of parameter optimization. The DT has easy-to-understand rule-based decision boundaries, whereas the RF, being a collection of many trees, minimizes overfitting and increases predictive stability.

Conversely, the ASD screening dataset is tabular, but CNN and DRNN were used to investigate if DL models could detect higher-order relationships between screening items and demographic variables. The CNN was not employed for image-based spatial learning, but rather the tabular feature vector was fed into one dimension as an input sequence, enabling the convolutional filters to learn local interactions between neighboring encoded features. The DRNN was also not designed to accurately simulate a real sequence of signals over time, but rather to determine if there's a dependency structure between successive responses in a screened sequence. This drawback, however, has been recognised and in some cases, CNN and DRNN may not be theoretically optimal for tabular datasets where there is no natural spatial or temporal structure. Thus, their inclusion is warranted empirically by comparing with the traditional ML models and the final MME-ASD ensemble.

The MME-ASD model has the advantage of having the diversity and complementarity of ML and DL paradigms. The ensemble method mitigates the flaws of individual learning algorithms, such as overfitting in DTs, parameter sensitivity in NNs, and constrained memory in CNNs, while inheriting the strengths of the others. Such a multi-model design will be used to ensure that the dynamic and static components of ASD-related features are well represented, resulting in a more accurate, stable, and generalizable diagnostic model for autism detection in adults.

2.2.1 Random Forest

Random Forest (RF) is an ensemble classification method, a supervised classification method that constructs an ensemble of independent DTs. Randomly selected samples and features are used to create each tree, and all the trees have a similar distribution. The final prediction is a combination of the individual tree's predictions. RF also takes into account the interactions between variables and uses the best split at each node to enhance classification accuracy [27]. A RF classifier is an aggregator of a family of classifiers called $h(x \mid \theta_1), h(x \mid \theta_2), \dots\ h(x \mid \theta_k)$, where k is the number of trees that have been selected at random from a model vector, and h is a classification tree, which means each θ_k has been selected at random from the vector of parameters. The algorithms for each tree will function in the same manner as those for conventional DTs, with the data being partitioned according to the value of the feature. When the results are aggregated in the manner given in Equation 1, the final output may be produced:

$$S(x) = \sum_{k} w_{\mathrm{k}}\, h_{\mathrm{k}}(x) \tag{1}$$

2.2.2 Decision Tree

Decision trees (DT)s are utilized for a variety of applications, including machine learning, image analysis, and pattern recognition, because they are effective for classification and decision-making. A DT is a collection of nodes connected together and arranged in a rooted tree where the top node is called the "root" and has no incoming edges. The mathematical model of the DT is shown below in Equation 2.

$$(x, y) = \left(x_1, x_{2,} x_3, \dots x_k, Y\right) \tag{2}$$

A DT may take the form of a leaf node with a class label or a test node with two or more children of a subtree. The test node will test the instance and lead the instance to a subtree for each possible outcome, depending on the attribute values. The classification process starts from the root node down to the other tree nodes. If the root is a test node, then the next subtree to be followed is determined by the outcome of the test. This continues until the leaf node is reached, in which case the leaf label is the prediction of the class of the instance.

The divide and conquer approach is a typical method for building DTs. If all the instances in a dataset are of the same class, the tree will turn into the leaf with the same class label. If not, an appropriate test is chosen which has the ability to give different results for at least two occasions. The data is then split based upon these results and the same process is repeated in each split.

Geometrically, a collection of x attributes is an x-dimensional feature space, and each instance is a point in this space. If we divide the instances based on the result of the test, it is the same as setting a decision boundary in this space. In many DT systems, there are tests where only one attribute, A, is used. For example, A = c could be tested for a nominal attribute, and x = t could be tested for a numerical attribute. These single-attribute tests create hyperplanes that are orthogonal to the attribute being tested. So, DTs using just one attribute test partition the feature space into hyperrectangular regions and each hyperrectangular region is linked to a class [28].

2.2.3 Neural Network

Neural networks (NN) are a type of computational model that is inspired by how the human brain learns and processes information. They are artificial units that are interconnected, and sometimes referred to as "neurons" or "parameters," which enable the model to learn patterns and adapt itself to new inputs. Every nerve cell has a single input, processes a function and generates an output. This output is then given as input to neurons in the next layer for further processing. This is repeated layer by layer until the information comes to the last layer, which generates the prediction or decision of the model.

An alternative way to view a NN is as a single, large function that takes inputs and produces an output. The intermediary processes performed by the multiple layers of neurons are typically unseen and happily automated. The mathematics behind them is both fascinating and intricate, and it merits more examination [29-31]. A nonlinear neuron is shown in Equation 3:

$$o = f\,(u) = f\,\left(\sum ni = 1\ w_i x_i\right) + b = f\,(w^T x + b) \tag{3}$$

2.2.4 Deep Recurrent Neural Networks

Deep Recurrent Neural Networks (DRNN) is based on the notion that a deep, hierarchical model may represent certain functions exponentially more effectively than a shallow one. Recent theoretical findings support this assumption. For instance, it demonstrated that a deep sum-product network may require exponentially fewer units than a shallow one to describe the same function. In addition, an abundance of empirical evidence supports this concept [32-34]. We assume that the same rationale should apply to RNN based on these results, as shown in Figure 2.

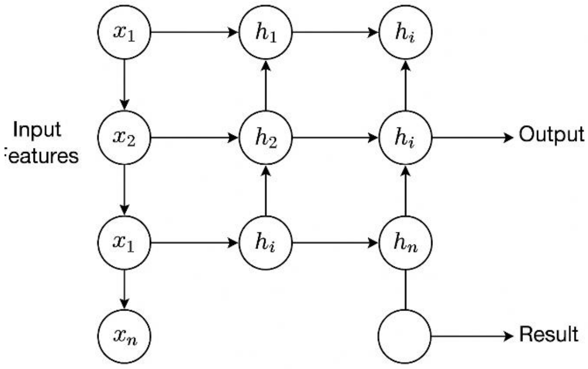


Fig. 2 *DRNN architecture*

According to Hermans and Schrauwen [35], DRNN, a form of RNN with several stacked hidden layers, is shown to naturally adopt a temporal hierarchy, with higher layers capturing interpretations aggregated over longer spans of text, while lower layers capture only short-term interactions between words. According to earlier studies, such hierarchies, when applied to sentences in natural languages, have the potential to better represent the multi-scale language effects characteristic of natural languages. As a result of the recent achievements of deep architectures in general and DRNN in particular, we investigate the possibility of using deep bidirectional RNNs, referred to as deep RNNs from here on, for the task of opinion expression extraction. We demonstrate that such models outperform traditional shallow (uni- and bidirectional) RNNs and earlier CRF-based state-of-the-art baselines, including the semiCRF model, in detecting DSE and ESE. This is the case for both types of errors.

2.2.5 Convolutional Neural Networks

A CNN is a type of NN with one or more Convolutional Layers. Other layers, such as nonlinear activation layers, pooling layers, and fully connected layers, can be added to create a deeper CNN architecture. While there are plenty of situations where CNNs are useful, there are also extra aspects to keep in mind during training. CNN convolutional filters are trained using backpropagation, with different structures depending on the specific task to be solved. In the case of face recognition, for instance, one filter might be trained to recognize edges, and the other might be trained to recognize patterns associated with eyes. Such filters are not manually created; instead, their values are learned during training [36].

Mathematically, a convolution is a function that is produced from two given functions via the process of integration. It illustrates how the shape of the second function influences the form of the first. The following is the convolution formula:

$$(f * g)(t) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(\tau) g(t-\tau) d\tau \tag{4}$$

In the convolutional layer, several filters are applied to the input data. This is done by taking the dot product of each filter element with each value from the input receptive field and then adding the sum of the results. The weighted sum is then used as part of the representation used for the next layer. There are three parameters that are most important for each convolution operation: Stride, Filter size, and Zero padding. The stride is the number of positions the filter is stepped per iteration. For instance, if stride=1, then the filter advances one step at a time prior to computing the value of the next output. The size of the filter (also known as the receptive field) should be the same for all filters in the same convolution. To size the original input matrix, zero padding is used to add rows and columns around the original input matrix [37].

The CNN architecture used in this work is shown in Figure 3. To prevent overfitting, our model consists of several layers that are each separated from the next by a dropout layer. These layers include a MaxPooling layer, a Flatten layer, two convolutional layers, and a Dropout layer. It is important to note that we used two convolutional layers that were separated by a dropout layer. This design has been proven to be enough and sufficient to provide the greatest results with nearly no overfitting difficulties, as will be shown and explained in the part devoted to the findings.

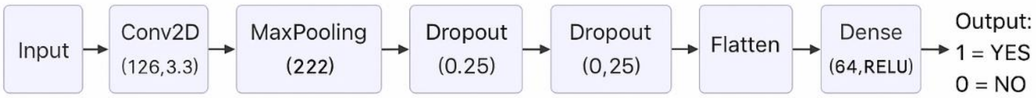


Fig. 3 *CNN architecture*

2.3 The MME-ASD Model

Ensemble learning is an appealing and powerful ML paradigm that fuses the predictions of several base learners to enhance the overall performance, generalization, and robustness. In contrast to relying on a single base learner, the ensemble architecture aggregates homogenous or heterogeneous learners, such as DT and NN, to capture comparable data instances. Such an approach minimizes variance and bias and often outperforms individual base learners across domains such as medical [38] and cybersecurity [39].

This work proposed a Multi-Model Ensemble for ASD Classification (MME-ASD) model. The MME-ASD is an ensemble model that combines the strengths of DL and traditional ML to achieve robust, precise predictions on the ASD dataset. As demonstrated in Figure 3, the current study comprises three main branches: traditional ML models, including RF, DT, and NN. The DL models, namely DRNN and CNN, and the proposed MME-ASD ensemble. For each branch, the models are trained independently on benchmark preprocessed ASD datasets to learn the dataset patterns. However, for the MME-ASD ensemble, models will produce a prediction probability, which in turn is fused through a Weighted Majority Voting (WMV).

WMV is a combination strategy that assigns the best weights to each model based on its evaluation metrics. This method reflects a hybrid ensemble strategy that enhances generalization, mitigates model bias, and guarantees elegant performance across diverse ASD data patterns. WMV is one of the most effective ensemble combination strategies, in which each model's output contributes to the final decision based on its binding reliability or weight. Hence, the final prediction is determined by adding the weighted votes of all learners, and the winning class is the one that records the highest score [40].

The mathematical model of the MME-ASD consists of five base models $k \in$ {RF, DT, NN, CNN, D RNN} that output a class label $h_{\mathrm{k}}(x) \in \{0, 1\}$ with a probability $p_{\mathrm{k}}(x) \in [0,1]$. The model processes x sample and produces the $y \in \{0, 1\}$ label. Within the process, it assigns non-negative weights w_{k} in which it results in a normalized $\sum_k w_{\mathrm{k}}$-1.

The predicted output of each k model is evaluated according to the WMV calculated score using:

$$S(x) = \sum_k w_{\mathrm{k}}\, h_{\mathrm{k}}(x) \tag{5}$$

The decision of a prediction run cycle by the five algorithms is made based on Equation 6.

$$\hat{y}(x) = \left[S(x) \geq \frac{1}{2} \sum_k w_{\mathrm{k}} \right] \tag{6}$$

Where the normalized weights, threshold is 0.5.

A validation-based practical weight setting assigns a voting weight to each model in the ensemble process, based on equation 7, which is proportional to the model's performance on a validation dataset. Once all the base models $k \in$ {RF, DT, NN, CNN, DRNN} are trained, their performance is checked with the help of an adopted performance measure like F1-score or accuracy. The products of the resultant values are then scaled to ensure that all weights are equal, so that stronger models would produce a greater contribution to the final decision in the ensemble as compared to weaker ones. This weighted approach to data helps make the ensemble more reliable and avoids cases where one model dominates the prediction.

$$w_{\mathrm{k}} = \frac{metric_{\mathrm{k}}}{\sum_j metric_j},\ metric_{\mathrm{k}} \in \text{evaluation metric of Section 2.4} \tag{7}$$

The following algorithm shows the complete processing steps of the proposed MME-ASD model. The MME-ASD algorithm consists of three major stages during the ensemble cycle. Every base learner $k \in$ (RF, DT, NN, CNN, DRNN) is trained on the training data individually and therefore captures various base learning patterns. Their validation performance is assessed using a reliability measure, F1-score, or accuracy, and those scores are normalized to weights wk, which assign each model a contribution to the final ensemble decision. During the Inference stage, a new sample x is provided, and each trained model gives a binary vote or probability output, a weighted sum $S(x) = \sum_k w_k h_k$ is calculated. The ensemble defines the sample as ASD when $S(x)$ is larger than some fixed parameter τ = 0.5 $\sum k\ w\ k$, which means that stronger models have a higher impact and weaker ones a smaller one. Lastly, at the management-per-stage level, the training time of an ensemble equals the sum of the training times of the five models, and the inference time increases linearly with the number of models, $O(|M|)$, even in the presence of multi-model aggregation.

Algorithm: MME-ASD

Input: Preprocessed ASD dataset D, base learners M = {RF, DT, NN, CNN, DRNN}, threshold τ = 0.5
Output: Final ASD class prediction y

1. Split D into training set Dtrain and testing set Dtest.
2. Split Dtrain into training subset Dtr and validation subset Dval.
3. For each model m in M do:
   3.1 Train m using Dtr.
   3.2 Predict validation labels using Dval.
   3.3 Compute validation performance score Sm.
4. Normalize model weights using:
   wm = Sm / ΣSm
5. Retrain each model m using the full Dtrain.
6. For each test sample x in Dtest do:
   6.1 Obtain prediction pm(x) from each model m.
   6.2 Compute weighted voting score:
   V(x) = Σ wm pm(x)
   6.3 If V(x) ≥ τ, assign y = ASD.
   6.4 Otherwise, assign y = Non-ASD.
7. Evaluate final predictions using accuracy, precision, recall, and F1-score.

2.4 Evaluation Metrics

The performance of the chosen classification algorithms used in the experiments is measured and compared using three evaluation metrics: error, accuracy, precision, and recall [8].

- Accuracy. The percentage of properly classified samples out of the total number of samples is called accuracy. The formula for calculating the accuracy is shown in Equation 8.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{8}$$

- **Precision**. Precision is defined as the percentage of correctly anticipated and computed samples, as shown in Equation 9.

$$Precision = \frac{(TP)}{(TP + FP)} \tag{9}$$

- **Recall.** True positive (TP) rate refers to the number of positive tuples that are correctly recognized when using recall or sensitivity. The formula for calculating the recall is shown in Equation 10.

$$Recall = \frac{(TP)}{(TP + FN)} \tag{10}$$

- **F1-score.** F1-score is a performance measure that is applied to determine the accuracy of a classification model, which is achieved by the combination of precision and recall into one value. It is especially applicable in situations where the dataset is uneven (i.e., one class is much smaller than the other, e.g., ASD positive cases). F1-score is the harmonic mean of precision and recall; therefore, a high F1-score indicates a low number of false positives and false negatives. The formula for calculating the F1-score is shown in Equation 11.

$$F1 = 2 \times \frac{(Precision + Recall)}{(Precision \times Recall)} \tag{11}$$

3. Results and Discussion

This section presents the experimental results obtained by applying both traditional and deep ML models, as well as the proposed MME-ASD ensemble classification framework. To assess models' stability, robustness, and generalization, different training-testing split ratios were employed. Benchmark evaluation metrics such as accuracy, precision, recall, and error rate are provided to highlight the strengths and weaknesses of each modeling approach. Moreover, the effect of the proposed ensemble on cross-validation is discussed, which reflects overall ensemble reliability and performance.

3.1 Results of ML and DL Models

This subsection presents the performance of individual traditional and deep ML models, namely, RF, DT, NN, CNN, and DRNN. Each of the later models was trained and evaluated on multiple train-test split ratios (i.e., 30-70, 40-60, 50-50, 60-40, and 70-30) to assess their ability to capture class-pattern relationships. Table 2 shows the performance for each model in different split ratios. Table 1 shows the performance comparison of traditional and DL models across different train-test splits. According to the results in Table 1, the 30-70 split appears to have yielded the most favorable outcomes. By way of comparison, it is quite easy to see that CNN has surpassed all other classifiers, and DRNN comes in second.

Table 1 *Initial results of individual ML and DL models*

Data split (%)	Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
30-70	RF	96.20	97	95	95.99
	DT	95.10	97	96	96.50
	NN	89.10	88	88	88.00
	CNN	99.53	98.39	100	99.19
	DRNN	99.06	98.43	99.33	98.88
40-60	RF	95.70	95	95	95.00
	DT	94.30	95	94	94.50
	NN	88.28	86	86	86.00
	CNN	98.58	97.6	99	98.30
	DRNN	97.16	96	97.28	96.64
50-50	RF	94.39	93	94	93.50
	DT	93.68	94	95	94.50
	NN	88.15	85	86	85.50
	CNN	97.73	96.65	97.79	97.22
	DRNN	97.16	96.46	96.46	96.46
60-40	RF	94.30	92	95	93.48
	DT	93.59	96	94	94.99
	NN	88.10	82	85	83.47
	CNN	98.35	97.8	98.1	97.95
	DRNN	97.64	97.1	97.1	97.10

	RF	94.20	91	98	94.37
	DT	93.50	93	92	92.50
70-30	NN	87.80	85	87	85.99
	CNN	96.55	95.4	96	95.70
	DRNN	96.96	95.95	96.55	96.25

The relatively small, structured data set may raise concerns about overfitting, given the high accuracy of the proposed MME-ASD model. To reduce this risk, the model was tested using a 60-40 split and 3-fold cross-validation. The result of the 3-fold cross-validation has also been presented with 99.57% accuracy, which is quite similar to the hold-out test result (99.65%). This means it is not the result of a specific favorable split. Moreover, to prevent data leakage, the processing parameters were used on the training data/training fold only. This is acknowledged as a weakness, as the current study did not test the model on an independent, external ASD dataset. In the future, the model will be tested externally with different ASD data sets from different populations to increase the model's generalizability. The CNN's confusion matrix for the split of 30/70 was selected as the best split and is shown in Figure 4. The CNN's performance is displayed in Figure 5. The confusion matrix provides direct counts of true positives, false positives, true negatives, and false negatives. The total of these variables equals 30% of the total 704 observations in the data set.

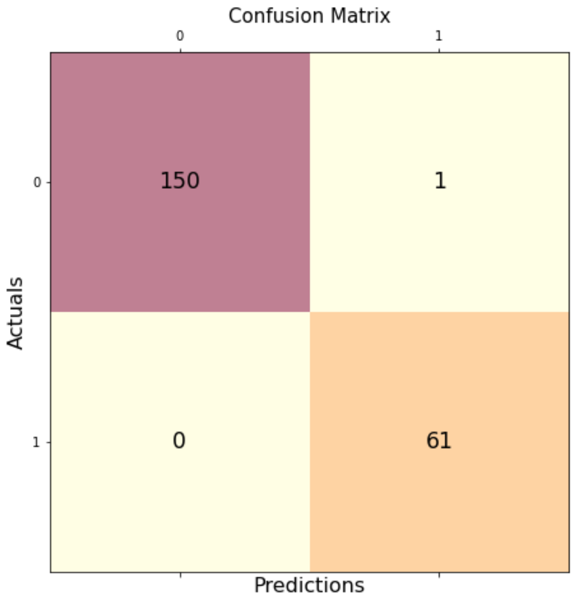
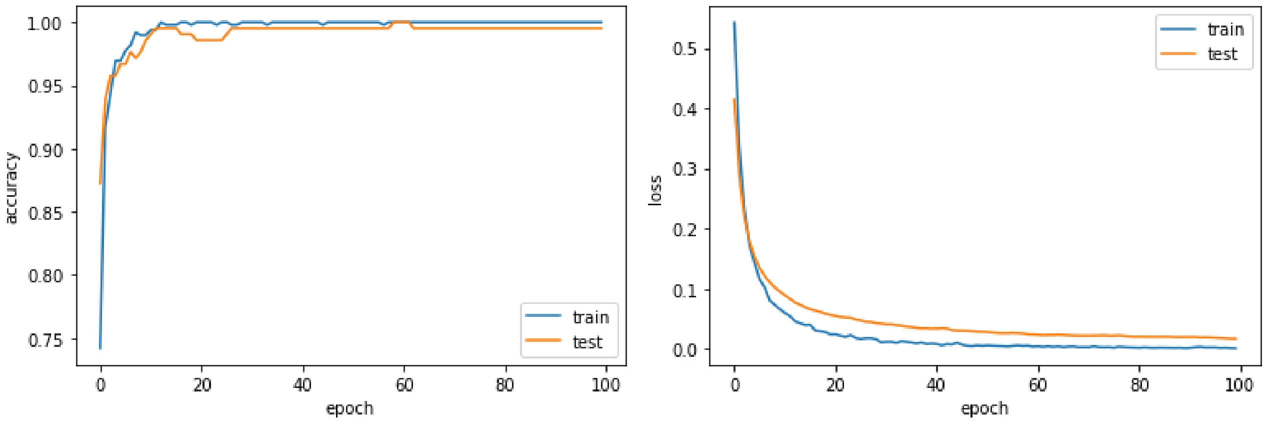


Fig. 4 *CNN confusion matrix*

A graphical depiction of the CNN accuracy and loss is shown in Figure 5. The figure shows the classifier's accuracy and loss (error) during both the training phase (blue line) and the testing phase (orange line). Compared with alternative models such as RF or DT, it is clear that CNN has achieved a very high level of performance while experiencing no overfitting whatsoever.



(a) CNN accuracy plot **(b) CNN loss plot**

Fig. 5 *CNN accuracy and loss*

3.2 Results of the MME-ASD Model

This subsection presents the results obtained by implementing the proposed MME-ASD ensemble model in Python using relevant ML and DL packages. The MME-ASD model design amalgamates the capabilities of traditional and DL models to get the best predictions of the ASD dataset. Firstly, the proposed MME-ASD was evaluated using a 60-40 training -testing split. Secondly, a 3-fold cross-validation was conducted to ensure robust performance prediction and minimize bias from random data partitioning. The selection of a 3-fold cross-validation was due to a compromise between computational complexity and accuracy in assessing performance, given that the proposed MME-ASD framework combines five different classifiers (ML and DL). The number of models needed for each fold would increase significantly with more-folds strategies like 5-fold or 10-fold cross-validation, thus adding significantly to computational complexity and run time. Furthermore, studies on ML for ASD disorders, including [24] and [37] highlighted effective validation processes without compromising the predictive assessment. Table 2 presents a performance comparison of the proposed MME-ASD across a training-testing split, 3-fold cross-validation, and the overall traditional and deep ML models. Table 2 shows the evaluation performance of the proposed MME-ASD ensemble using a 60–40 train–test split and 3-fold cross-validation, compared with traditional and DL models.

Table 2 *The results of the MME-ASD and baseline models*

Evaluation method	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
RF	94.9	93.6	95.4	94.49
DT	94	95	94.2	94.60
NN	88.3	85.2	86.4	85.80
CNN	98.14	97.4	98	97.70
DRNN	97.6	96.8	97.3	97.05
MME-ASD 60-40 split	99.65	98.70	100	99.35
MME-ASD 3-fold cross-validation	99.57	99.48	98.94	99.21

Table 2 shows that under a 60-40 split, the proposed MME-ASD ensemble achieved superior accuracy (99.65%), recall (1.0), precision (98.70%), and F1-score (99.21%). These values indicate that the MME-ASD model performs excellently on the held-out test dataset, achieving perfect recall, suggesting that all positive patterns in the test dataset are correctly identified.

On the other hand, when the proposed ensemble model was evaluated on a 3-fold cross-validation, it performed well, with consistent results, achieving high evaluation metrics of 99.57%, 99.48%, 99.94%, and 99.21% for accuracy, precision, recall, and F1-score, respectively. There is a slight reduction in the results compared with a 60-40 split, and this reduction is expected, as the cross-validation method provides a more general and comprehensive estimate of the model's behavior by testing the model on multiple folds of the dataset. Nevertheless, the drop in performance was minimal, less than 0.1% in accuracy, corroborating that the model is stable and not overfitting. Moreover, Table 2 shows the overall performance of traditional and deep ML models. Results show that both CNN and DRNN achieved high accuracy metrics of 98.14% and 97.6%, respectively, outperforming traditional ML models like RF and DT. As the NN represents a conceptually high-performance model, it achieves a relatively lower 99.3% accuracy, perhaps due to its limitations in depth and capacity compared with CNN and DRNN models.

The confusion matrix of the MME-ASD model with the 30/70 data split (optimal split for CNN) is shown in figure 6. The confusion matrix is a chart that displays the numbers of true positives, false positives, true negatives and false negatives. These outcomes add up to 30% of all 704 outcomes in the data set.

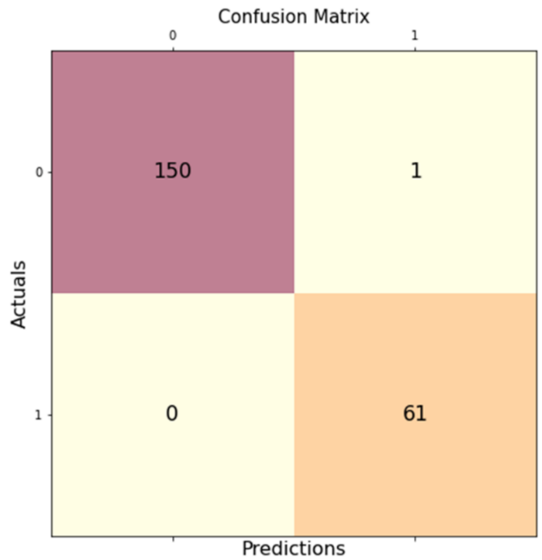
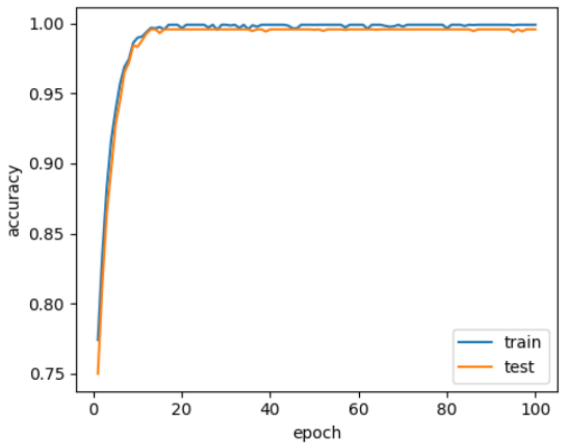
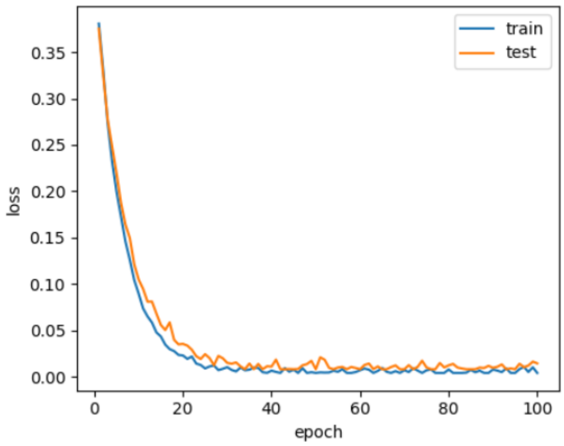


Fig. 6 *MME-ASD confusion matrix*

Figure 7 shows the MME-ASD classifier's accuracy and loss (error) during training (blue line) and testing (orange line). Compared with alternative models such as RF, DT, or CNN, the MME-ASD model has achieved a very high level of performance with no overfitting.



(a) CNN accuracy plot



(b) CNN loss plot

Fig. 7 *MME-ASD accuracy and loss*

The profound observations demonstrate that the DL architecture can capture sophisticated feature representations compared to traditional ML models, resulting in superior accuracy and recall. However, compared with our proposed MME-ASD ensemble, DRNN and CNN were recording high performance by a significant margin, with the MME-ASD ensemble achieving over 99.5% accuracy with near-perfect recall. These results indicate that the proposed MME-ASD is efficiently leveraging the complementary strengths of the constituted learners, fusing the generalization ability of traditional ML models with the high-performance learning of deep models.

Ensemble-based techniques for the classification of ASD have been studied in the past. For instance, Arunkumar and Surendran [40] used an ensemble ML and max-voting approach for the diagnosis of ASD, and Zhang-James et al. [41] used an ensemble classification approach using neuroimaging features for the diagnosis of ASD. The deep ensemble model and transformer-based ensemble model for ASD prediction have also been explored in recent studies, using facial images and health registry data. Most of the current literature is limited to either traditional ML or DL-based approaches, or to particular types of data, such as MRI, facial images, or registry data. This has been contrasted with the proposed MME-ASD model, which fuses classifiers from traditional and DL techniques by employing weighted majority voting on structured ASD screening data to enhance the stability and generalizability of the model's classification. Table 3 shows the higher classification accuracy rate of the proposed MME-ASD framework than several recent ASD classification studies. The results showed that using a multi-model ensemble based on the validation process (weighted majority voting) based on heterogeneous models (ML and DL) enhances predictive robustness and classification accuracy on structured ASD screening data.

Table 3 *The results of the MME-ASD and state-of-the-art models*

Study	Method	Dataset/Input	Best Accuracy (%)
Vaishali and Sasikala [22]	ML with feature selection	ASD behavioral dataset	97.95
Kosmicki et al. [23]	ML with ADOS behavioral features	ASD behavioral data	98.27
Mashudi et al. [7]	ML classification approach	Adult ASD dataset	96.00
Luongo et al. [24]	ML classifiers with motion analysis	Kinematic ASD data	87.30
Twala and Molloy [27]	Ensemble classifiers	ASD therapy prediction	97.00
Proposed MME-ASD	RF + DT + NN + CNN + DRNN with WMV	Adult ASD screening dataset	99.65

The results of the proposed MME-ASD model were compared to the previous studies of ASD classification as further proof of the effectiveness of the model. Using ML with feature selection, Vaishali and Sasikala [22] obtained a range of accuracy from 92.12% to 97.95%. Selected behavioral features from ADOS modules were used by Kosmicki et al. [23] to get 98.27% and 97.66% accuracy. The value of combining classifiers was also highlighted by Twala and Molloy [27], who investigated ensemble-based ASD prediction. By comparison to these studies, the proposed MME-ASD was the most accurate model with a 99.65% accuracy for a 60-40 split and 99.57% accuracy for a 3-fold cross-validation. This improvement suggests that the performance of the classification of ASD can be improved using traditional ML and DL models by combining them based on weighted majority voting.

4. Conclusions and Future Work

Autism spectrum disorder (ASD) is a common neurodevelopmental illness that is defined by widespread issues throughout early infancy, including limited, repetitive interests and behaviors. This study proposed a new ensemble approach, MME-ASD, that integrates both traditional and deep ML models for ASD classification, using the CRISP-DM framework to achieve the objective. The constituent classifiers of the proposed MME-ASD ensemble are RF, DT, NN, DRNN, and CNN, which are combined using a weighted majority voting framework. To assess the proposed ensemble, three diverse experiments have been conducted; the first uses different train-test splits with both traditional and deep ML models as standalone classifiers. The second implements the proposed MME-ASD ensemble of a 60-40 split. The last was implemented on a 3-fold cross-validation. The experimental results show the superiority of the proposed MME-ASD compared with both traditional and deep ML models, in which the MME-ASD achieved the best performance of 99.65%, 98.70%, and 99.35% for accuracy, precision, and F1-score, respectively, and perfect recall under a 60-40 split. Also, it shows high performance under 3-fold cross-validation, achieving 99.57% accuracy. These results demonstrated that the proposed model exhibits robust classification behavior with stable outputs across different conditions. In the near future, authors will focus on enhancing model interpretability and minimizing computational overhead for real-time testing. Moreover, new learning mechanisms will be explored to further improve prediction on the ASD dataset.

Acknowledgement

The authors would like to thank the University of Mosul for its support.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of the paper.

J. of Soft Computing and Data Mining Vol. 7 No. 2 (2026) p. 371-387

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Dhafar Fakhry Hasan, Maha A. Abdul-Jabar;* ***data collection:*** *Mawadah Mohammed Suliman;* ***analysis and interpretation of results:*** *Massila Kamalrudin;* ***draft manuscript preparation:*** *Dhafar Fakhry Hasan, Maha A. Abdul-Jabar, Mawadah Mohammed Suliman, Massila Kamalrudin. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Khudhur, D. D., & Khudhur, S. D. (2023). The classification of autism spectrum disorder by machine learning methods on multiple datasets for four age groups. Measurement: Sensors, 27, 100774.

[2] Thabtah, F. (2019). Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. Informatics for Health and Social Care, 44(3), 278–297. https://doi.org/10.1080/17538157.2018.1433677

[3] Alcañiz Raya, M., Marín-Morales, J., Minissi, M. E., Teruel Garcia, G., Abad, L., & Chicchi Giglioli, I. A. (2020). Machine learning and virtual reality on body movements' behaviors to classify children with autism spectrum disorder. Journal of Clinical Medicine, 9(5), 1260. https://doi.org/10.3390/jcm9051260

[4] SARWANI, I. S., & BHASKARI, D. L. (2024). A novel ensemble meta-features integration technique for autism spectrum disorder detection. Journal of Theoretical and Applied Information Technology, 102(10).

[5] Tang, Y., Tong, G., Xiong, X., Zhang, C., Zhang, H., & Yang, Y. (2023). Multi-site diagnostic classification of Autism spectrum disorder using adversarial deep learning on resting-state fMRI. Biomedical Signal Processing and Control, 85, 104892.

[6] Buffle, P., & Naranjo, D. (2021). Early identification and diagnosis of autism spectrum disorder: a literature review on evidence-based recommendations. Rev Ecuatori Ped, 22(3), 1-19.

[7] Mashudi, N. A., Ahmad, N., & Noor, N. M. (2021). Classification of adult autistic spectrum disorder using machine learning approach. IAES International Journal of Artificial Intelligence, 10(3), 743.

[8] Squarcina, L., Nosari, G., Marin, R., Castellani, U., Bellani, M., Bonivento, C., ... & Brambilla, P. (2021). Automatic classification of autism spectrum disorder in children using cortical thickness and support vector machine. Brain and behavior, 11(8), e2238.

[9] Saleh, A. Y., & Chern, L. H. (2021). Autism Spectrum Disorder Classification Using Deep Learning. International journal of online & biomedical engineering, 17(8).

[10] Rahman, M. A., Hossain, M. M., Singh, S. P., & Sharmin, N. (2025). Predicting early ASD traits of adults and toddlers using machine learning and deep learning with explainable AI and optimization. Neural Computing and Applications, 1-28.

[11] Suk, H.-I., Lee, S.-W., & Shen, D. (2015). Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. Brain Structure and Function, 220(2), 841–859. https://doi.org/10.1007/s00429-013-0687-3

[12] Liu, S., Liu, S., Cai, W., Pujol, S., Kikinis, R., & Feng, D. (2014). Early diagnosis of Alzheimer's disease with deep learning. In 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI) (pp. 1015–1018). IEEE. https://doi.org/10.1109/ISBI.2014.6868045

[13] Kim, J., Calhoun, V. D., Shim, E., & Lee, J.-H. (2016). Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. NeuroImage, 124, 127–146. https://doi.org/10.1016/j.neuroimage.2015.05.018

[14] Wing, L. (1997). The autistic spectrum. The Lancet, 350(9093), 1761–1766. https://doi.org/10.1016/S0140-6736(97)09218-0

[15] Lakkapragada, A., Kline, A., Mutlu, O. C., Paskov, K., Chrisman, B., Stockham, N., ... & Wall, D. P. (2022). The classification of abnormal hand movement to aid in autism detection: Machine learning study. JMIR Biomedical Engineering, 7(1), e33771.

[16] Jiao, Y., Chen, R., Ke, X., Chu, K., Lu, Z., & Herskovits, E. H. (2010). Predictive models of autism spectrum disorder based on brain regional cortical thickness. NeuroImage, 50(2), 589–599. https://doi.org/10.1016/j.neuroimage.2009.12.047

[17] Zhang, F., Chen, G., He, Y., & Liu, Y. (2021). Identifying autism spectrum disorder subtypes using multi-site structural MRI with multivariate pattern analysis. NeuroImage: Clinical, 30, 102610. https://doi.org/10.1016/j.nicl.2021.102610

[18] Li, W., Wang, Z., Zhang, L., Qiu, S., & Shen, D. (2022). Automated diagnosis of autism spectrum disorder using multi-feature structural MRI and deep learning. Human Brain Mapping, 43(6), 1975–1990. https://doi.org/10.1002/hbm.25759

[19] Yao, D., Sui, J., Wang, M., Jia, Y., & Calhoun, V. D. (2023). Gray matter alterations in children with autism spectrum disorder: A multi-site voxel-based morphometry and machine learning study. NeuroImage: Clinical, 38, 103397. https://doi.org/10.1016/j.nicl.2023.103397

[20] Thabtah, F. (2017). Autism spectrum disorder screening: Machine learning adaptation and DSM-5 fulfillment. In Proceedings of the 1st International Conference on Medical and Health Informatics (pp. 1–6). ACM. https://doi.org/10.1145/3107514.3107517

[21] Mythili, M. S., & Shanavas, A. M. (2014). A study on autism spectrum disorders using classification techniques. International Journal of Soft Computing and Engineering, 4(5), 88–91.

[22] Vaishali, R., & Sasikala, R. (2018). A machine learning-based approach to classify autism with optimum behaviour sets. International Journal of Engineering and Technology, 7(4), 18.

[23] Kosmicki, J. A., Sochat, V., Duda, M., & Wall, D. P. (2015). Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. Translational Psychiatry, 5(2), e514. https://doi.org/10.1038/tp.2015.9

[24] Luongo, M., Simeoli, R., Marocco, D., Milano, N., & Ponticorvo, M. (2024). Enhancing early autism diagnosis through machine learning: Exploring raw motion data for classification. PLoS One, 19(4), e0302238.

[25] Nagamani, C., Devi, T. R., Babu, G. G., & Sindhu, V. (2024, April). Comparative exploration of machine learning models for enhanced autism detection. In 2024 International Conference on Inventive Computation Technologies (ICICT) (pp. 925-932). IEEE.

[26] Biju, T. P., Thanka, M. R., Edwin, E. B., Ebenezer, V., Kirubakaran, S., & Jenefa, A. (2024, June). Neural networks unveiling autism: Deep learning approaches for spectrum detection. In 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) (pp. 445-450). IEEE.

[27] Twala, B., & Molloy, E. (2023). On effectively predicting autism spectrum disorder therapy using an ensemble of classifiers. *Scientific Reports*, *13*(1), 19957.

[28] Ayyadurai, M., Sujatha, K., Deeptha, R., Preethi, D., & Dinesh, M. (2024, June). Exploring Data Mining Techniques for Early Autism Detection using Random Forest Algorithm. In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.

[29] Ahmed, M., Hussain, S., Ali, F., Gárate-Escamilla, A. K., Amaya, I., Ochoa-Ruiz, G., & Ortiz-Bayliss, J. C. (2025). Summarizing recent developments on autism spectrum disorder detection and classification through machine learning and deep learning techniques. Applied Sciences, 15(14), 8056.

[30] Gulcehre, C., Cho, K., Pascanu, R., & Bengio, Y. (2014). Learned-norm pooling for deep feedforward and recurrent neural networks. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 530–546). Springer.

[31] Delalleau, O., & Bengio, Y. (2011). Shallow vs. deep sum-product networks. Advances in Neural Information Processing Systems, 24, 666–674.

[32] Ganesh, K., Umapathy, S., & Thanaraj Krishnan, P. (2021). Deep learning techniques for automated detection of autism spectrum disorder based on thermal imaging. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 235(10), 1113-1127.

[33] Aljuaid, A., Althagafi, A., & Thafar, M. A. (2025). Transfer Learning with Transformer-Based Models and Explainable AI for Autism Detection Using Brain MRI.

[34] Cheekaty, S., & Muneeswari, G. (2024). Enhanced multilevel autism classification for children using eye-tracking and hybrid CNN-RNN deep learning models. Neural Computing and Applications, 1-24.

[35] Alsubai, S. (2025). Deep Learning-Based Matrix of Mind Multi-Diagnostic Model for Identifying Dementia, Autism Spectrum Disorder, and Learning Disabilities. Journal of Disability Research, 4(5), 20250691.

[36] El-Shafai, W., & Algarni, A. D. (2024). Predicting of Uncertainty Facial Expressions in Children with Autism Spectrum Disorder Based on an Improved Fuzzy Convolutional Neural Network Model.

J. of Soft Computing and Data Mining Vol. 7 No. 2 (2026) p. 371-387

[37] Mostafa, S. A., Hasan, D. F., Abdul-Jabar, M. A., & Mustapha, A. (2025). An ensemble learning model for multi-type cancer prediction in clinical diagnostic decision support systems. Journal of Soft Computing and Data Mining, 6(1), 230–246.

[38] Al-Shakarchi, A. H., Al-A'araji, N. H., & Al-Mamory, S. O. (2023). A novel ensemble method for network-based anomaly intrusion detection system. In A. M. Al-Bakry, S. O. Al-Mamory, M. A. Sahib, L. E. George, J. A. Aldhaibani, H. S. Hasan, & G. S. Oreku (Eds.), New trends in information and communications technology applications (pp. 191–229). Springer Nature. https://doi.org/10.1007/978-3-031-33414-4_15

[39] Alshammeri, M., Tariq, N., & Humayun, M. (2025). Ensemble Deep Neural Networks for Reliable Autism Diagnosis in Children: A Performance and Interpretability Study. Journal of Disability Research, 4(5), 20250671.

[40] Arunkumar, A., & Surendran, D. (2022). Autism Spectrum Disorder Diagnosis Using Ensemble ML and Max Voting Techniques. *Computer Systems Science & Engineering*, *41*(1).

[41] Zhang-James, Y., Buitelaar, J. K., ENIGMA-ASD Working Group, van Rooij, D., & Faraone, S. V. (2021). Ensemble classification of autism spectrum disorder using structural magnetic resonance imaging features. *JCPP advances*, *1*(3), e12042.