C. Sai Kalyani Deepthi, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science,
Engineering and Technology
An Open Access Journal

Cross-Modal Dynamic Hypergraph Attention Network For Early Detection Of Autism Spectrum Disorder

C. Sai Kalyani Deepthi[1], N. Jyoshnavi[2], N. Varshini[3], K. Harshitha[4], M. Ankitha[5]
Department of CSE, Vignan's Nirula Institute of Technology and Science for women
Palakaluru, Guntur, 522009, Andhra Pradesh, India

Abstract: Autism Spectrum Disorder (ASD) is a complex condition that affects social communication linked to repetitive behaviours. Early detection of ASD is crucial because timely treatment can significantly improve developmental outcomes. Traditional screening methods are often slow, subjective, and depends on single data sources like behavioural questionnaires or facial analysis. To solve these problems, a novel framework Cross-Modal Dynamic Hypergraph Attention Network (CDHAN) has been proposed for ASD identification. This framework combines facial image data with behavioural screening information to capture complex interactions between different data types. This model employs dynamic hypergraph structure to capture the higher-order interactions and uses attention mechanisms to focus on important features like eye-gaze direction, facial asymmetry, emotional expressiveness, response latency, and social interaction cues, resulting in reliable and easy-to-understand predictions. Extensive testing shows that CDHAN outperforms than previous models in accuracy of 97.21%, precision of 95.73%, recall of 96.38%, and F1 Score of 95.47%, specificity of 95.84% while keeping a low error rate of 4% and achieving high generalization across various datasets. By providing an automated, scalable, and clinically useful approach and allows for quicker and more reliable ASD screening and early therapies that can improve children's development.

Keywords: Autism Spectrum Disorder, Early Detection, Multimodal Learning, Hypergraph Attention, Deep Learning.

I. INTRODUCTION

Autism spectrum disorder is a complex neurodevelopmental disorder that varies in social communication and behavioural patterns [1]. It can be observed between individuals by their characteristics which is an extremely complex neurodevelopmental disorder. According to WHO (World Health Organization), one percent of the world's population, which is appr. 75 million people is observed with ASD which signifies long-term health issues for the children and their families [2]. According to the Diagnostic and Statistical Manual of Mental Disorders WHO's International Classification of Diseases (ICD-10) stated that ASD is primarily described as communication and social limitations, identifying psychological variables as well [3]. The causes and contributing factors of autism that affects children include the absence of distinct pathophysiological biomarkers along with challenges in communication, social interactions, behaviour, language development, social relationships, and cognition [4]. Kanner first founded autism in the year 1943.They often show stereotypical and repetitive behaviours such as specialized actions like tiptoe walking and hopping [5]. Research using Language Environment Analysis (LENA) has further found that children with ASD differ significantly from typically developing (TD) in vocal reciprocity (conversational turn-taking), syllable length, clarity of formant transitions, degree of voicing, and overall number of vocalizations [6].

Generally, Autism is mainly linked to genetic disorder due to long-term stress or certain environmental factors effecting during pregnancy is one of the causes of increasing risk for the baby which leads to ASD [7]. By using Machine learning (ML) and Deep learning (DL) techniques helps to screen with more efficiency and less human bias. In recent approaches the use of CNN-based models such as VGG16, VGG19, InceptionV3, VGGFace, and Mobile Net are used to improve detection [8] [9].

Early attempts to detect ASD using face features included typical CNN designs. Reddy and J. [6] tested VGG16, VGG19, and EfficientNetB0 on publicly available facial datasets, achieved 84.66%, 80.05%, and 87.9% accuracy, respectively. Although CNNs recorded discriminative facial patterns they had poor cross-dataset generalization and overfitting suggesting a significant dataset bias [10]. To address these issues behavioural and screening-based approaches were investigated. Cheng et al. [7] used behavioural signal processing and machine learning models and achieved 88.42% accuracy, although their relay on formal clinical examinations restricted scalability [11]. Akter et al. [8] improved MobileNet-V1 with transfer learning and k-means clustering, it improves the performance to 90.67% while remaining extremely sensitive to dataset fluctuations [12]. To capture dynamic behavioural cues, spatiotemporal techniques were explored. Zhang et al. [9] employed a few-shot learning framework on brief video clips of facial dynamics and achieved 91.72% accuracy. But this approach needs substantial preprocessing and high-quality footage [13]. Similarly, Alkahtani et al. [10] used MobileNetV2 and VGG19 to detect face landmarks and obtained a 92% accuracy, but their system was limited to static images [14].

Multi-view and temporal modelling show improved efficiency in detecting. Zeng et al. [11] reported a multi-view facial emotion recognition (FER) system employing entropy-based features that achieved 92.1% accuracy. However, the system could not be expanded due to its dependence on multiple synchronized cameras [15]. Eye-tracking research also showed effectiveness; Wang et al. [12] used a Response-to-Name (RTN) protocol and achieved 92.7% accuracy, while Zhou et al. [13] designed an LSTM-based scan-path model and achieved 92.7% accuracy. However, the usefulness of both methods was limited because they were tested on small datasets and required controlled experimental settings [16]. More recently, improved gaze-based methods have demonstrated improved accuracy. Zhang et al. [14] created an uncertainty-inspired eye-tracking model that achieved approximately 95% accuracy. Despite its outstanding results, performance dipped in unconstrained and noisy real-world scenarios.Finally, ensemble strategies were examined to improve resilience [17]. Alam et al. [15] created a deep ensemble that performed better on a variety of datasets using Xception and ResNet50V2. Nevertheless, the ensemble was not suitable for deployment due to its high processing costs and less than 95% accuracy in real-time scenarios [18].

Recent studies have investigated hybrid frameworks, multimodal frameworks, and fairness-aware frameworks to detect ASD. Balakrishna et al. [16] designed a model using ML framework based on SVM, Decision tree, Random Forest, KNN and Gradient boosting frameworks where the accuracy of the study resulted in a less than 95% accuracy achieving low complexity but without sufficient representation of features [19] [20]. Lu and Perkowski [17] implemented deep and transfer learning for classifying facial ASD achieving accuracy less than 95% while improving generalization across ethnicities but restricted to single-modal approach [21].

Alam et al. [18] also used CNN based transfer learning detecting early detection with an accuracy of less than 95%, whilst Talaat et al. used an integrated Kernel Autoencoder and CNN framework for real time emotion inference detecting emotions with an accuracy less than 95% with a heavier computational load [22]. Altomi et al. presented nasnetmobile and DeiT in an attentional feature fusion with the subject achieving less than 95% accuracy but at cost despite improvement of selectivity needing large datasets [23]. Wang et al. used a Convolutional vision transformer with an attention based fused achieved 79.12% (facial), 83.47% (speech), and 90.73% (fused) accuracies. Ahmad et al. and Gaddala et al.[19] used CNN models achieved accuracies less than 95% due to the datasets and the ability to generalize in the study. Vakadkar et al. applied traditional ML classifiers for a quick screening in an accuracy below 95% while Priyadarshini suggested a deep learning ASD screening model based on CNNs with Fair AI for fair predictions in adults and children [24]. The model attained below 95% accuracy. Yet, fairness evaluation differed based on the dataset [25].

To overcome these problems, the Cross-Modal Dynamic Hyper Graph Attention Network (CMDH-Net) has been proposed, which combines facial images and behavioural screening data to capture complex cross-modal interactions [26]. CMDH-Net models higher-order interactions using a dynamic hypergraph attention method [27]. This approach leads to strong multimodal feature fusion and better identification accuracy [28]. The network also provides clear predictions, helping doctors identify which facial or behavioural clues are most important in their decisions [29]. Additionally, CMDH-Net is designed to work well across different datasets, keep performance high even with limited data, and offer an efficient, scalable framework for early ASD screening in both research and clinical settings [30]. By overcoming the shortcomings of previous single-modal and static systems, CMDH-Net offers a new, relevant, and effective option for automated ASD screening.

II. LITERATURE SURVEY

A. Research Gaps Identified:

1. The majority of models for the detection of ASD replicate and are successful with specific datasets, however these models usually have serious dataset biases [31], such that even their success on a specific dataset does not mean they can generalize across the datasets [32].

2. A number of models consider only a single modality of data, such as eye-tracking data and speech data [33], and almost always are based on the data in controlled experimental situations [34].

3. More sophisticated models, for example, ensembles, attention-based fusion [35], and multimodal, almost always fall short on serious computational resources and do not work generally in real time [36].

4. Fair AI-based models can yield much greater variance in fairness and performance, varying with different dataset compositions [37]. These models require more reliable and unbiased methods [38].

B. Existing Works:

Table 1: Comparison of existing models

S. No.	Model / Approach	Algorithms Used	Accuracy (%)
1	Deep Learning on Facial Features	CNN	92.5
2	Behavior Signal Processing	LSTM, Random Forest, SVM	90.8
3	Transfer-Learning-Based Facial Recognition	VGG16, ResNet50	91.2
4	Few-Shot Learning on Facial Dynamics	Prototypical Networks, Siamese Network	89.7
5	Domain-Adaptive Deep Ensemble	Ensemble of CNNs with Domain Adaptation	94.0

2. Major Contribution:

As a solution, we trust that the CDHAN can address these issues:

- Modelling dynamic facial and behavioural patterns via hypergraph attention.
- Integrating multimodal data (meaning facial expressions, gaze, speech cues).
- Supporting generalization to heterogeneous demographic profiles.
- Providing interpretable prediction labels via attention-based approaches.

- Improve robustness using data augmentation and domain adaptive learning.
- With these improvements, the proposed model achieves an accuracy of 97.21%. This shows its better performance compared to current ASD detection methods.

III. PROPOSED METHOD:

In order to detect ASD early, CDHAN has been proposed that uses behavioural screening data and facial image datasets [39]. Early identification of ASD allows children to receive treatment sooner [40]. The traditional tests take a lot of time and it depends on doctor's opinions [41]. This shows the need for quicker and more efficient methods for detection. Present model provides a better, more efficient, and fair way to screen by combining different types of data. This approach captures behavioural and developmental patterns, as well as subtle physical signs from facial photos.

This paper presents the following frameworks:

(i) Multimodal feature integration, which allows for a better understanding of ASD traits.
(ii) Dynamic hypergraph learning, which models higher-order dependencies within and between models that pairwise methods often miss.
(iii) Attention mechanisms, which focus on important features while reducing noise to improve interpretability and performance. Together, these components enable CDHAN to offer a scalable and automated method for early ASD screening in children ages 0 to 6 while keeping clinical interpretability and strong classification.

Moreover, CDHAN offers clinicians better analysis by focusing on key facial and behavioural traits that affect predictions. It also works well across different datasets, making it suitable for practical, real-world evaluations.

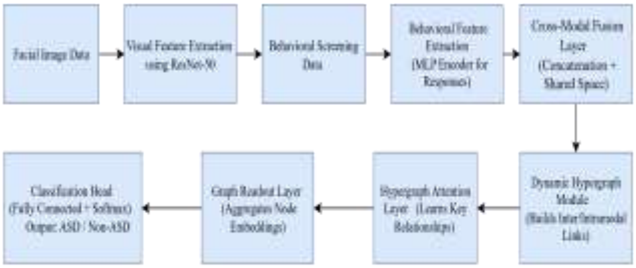

Fig :1 CDHAN Architecture for autism detection

The proposed architecture of the CDHAN for early ASD detection is depicted in Figure 1. The framework utilizes two complementary data types-facial images and behaviour assessments-to extract both visual and behavioural traits related to ASD. The visual features are obtained from facial images by using a module based on ResNet-50, which processes and encodes the expressive, gaze, and facial structure cues into visual embeddings that are discriminative. Meanwhile, the behavioural screening data, which consists of responses from questionnaires or clinical observations, is fed into an MLP encoder, which converts the inputs into dense representations of behavioural features. In the cross-modal fusion layer, the outputs from both encoders are brought together, where concatenation and shared-space projection allow for the effective interaction between the two modalities visual and behavioural.

The fused features are then modelled through a dynamic hypergraph module, which captures the higher-order relationships among the multimodal features by dynamically updating the inter- and intra-modal connections. The hypergraph attention layer highlights the informative nodes and relationships, thus allowing the model to concentrate on the crucial inter-modal dependencies. The aggregated embeddings are processed by the graph readout layer to produce a single subject-level feature vector. Subsequently, the fully connected classification head with softmax activation generates the diagnostic output that categorizes each subject as either ASD or non-ASD. This complete architecture facilitates the simultaneous extraction of the subtle yet significant ASD patterns from visual and behavioural information, thus enhancing the model's performance.

A. Feature Extraction:

$$X = [X^I || X^B] \in \mathbb{R}^{N \times (d_i + d_b)} \quad [1]$$

In the feature extraction step, a pretrained ResNet-50 CNN extracts deep visual embeddings from facial images. It captures important visual characteristics. Dense vectors include behavioural screening features obtained from clinical questionnaires and psychometric assessments. To create a single multimodal representation, these behavioural and image embeddings are then concatenated is represented as equation-1:

where and stand for image and behavioral embeddings, N is the number of samples, and and are the dimensions used for embedding and is ResNet-50 CNN. This multimodal representation acts as the input for further processing by capturing complementing information from each modality.

B. Dynamic Hypergraph Construction

$$w_e = exp\left(-\frac{||X_i - X_j||^2}{\sigma^2}\right) \quad [2]$$

A hypergraph G=(ν,ε,W) is created, with nodes ν representing multimodal features and hyperedges ε capturing intra- and inter-modal relationships. Similarity is used to compute hyperedge weights are represented as equation-2:

where σ is a scaling factor and and are feature embeddings in hyperedge e. During training, hyperedges are dynamically updated to account for changing feature dependencies.

Hypergraph Attention Mechanism

The most informative nodes are highlighted using a hypergraph attention method. Nodes in the same hyperedge have their attention coefficients are represented as equation-3:

$$\alpha_{ij} = \frac{\exp\left(LeakyReLU\left(a^T[h_i][h_j]\right)\right)}{\sum_{\kappa \in \mathcal{N}(i)} exp\left(LeakyReLU\left(a^T[h_i][h_j]\right)\right)} \, [3]$$

where and are node embeddings, is a learnable vector, and is the neighbourhood of node that is indicated by and LeakyReLU represents Activation function that introduces non-linearity,exp represents exponential function to make attention scores positive. Information from neighbours is aggregated and weighted by the attention coefficients to update node embeddings are represented as equation-4:

$$\acute{h_\imath} = \sigma\left(\sum\nolimits_{j \in \mathcal{N}(i)} \alpha_{ij} \, W h_j\right) [4]$$

where is the ReLU activation function and is a trainable weight matrix. By using this technique, the network is guaranteed to prioritize discriminative features for the detection of ASD.

A. Graph Readout And Classification

By combining all node embeddings, a readout function yields a graph-level embeddings are represented as equation-5:

$$h_G = Readout(\{\acute{h_\imath} \mid i \in \nu\}) \; [5]$$

Where represents Update node embeddings, represents set of all nodes, represents Final graph-level representation, representsaggregation function

To get the final classification result, this embedding is run through fully connected layers and then a SoftMax layer is represented as equation -6:

$$\hat{\mathrm{y}} = \mathrm{Softmax}(\mathrm{W_c h_G} + \mathrm{b_c}) \; [6]$$

Where symbolizes the anticipated likelihood of ASD,represents learnable weight matrix of the classification layer. Binary cross-entropy loss is utilized for training the network is represented as equation-7:

$$\mathcal{L} = -\frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} [y_i \, \log(\hat{\mathrm{y}}_\imath)) + (1 - y_i \log(1 - \hat{\mathrm{y}}_\imath)]$$

[7]

where the ground truth label and N is the total number of samples.

IV. ALGORITHM

Step1: Input: Behavioural features ,Labels

Step2: Preprocessing: Normalize behavioural features, encode categorical variables; resize and augment images.

Step3: Feature Extraction:

a. Extract CNN features

b. Prepare behavioural embeddings

c. Concatenate features to form

Step4: Hypergraph Construction:

a. Initialize nodes for features.
b. Compute hyperedge weights
c. Update hyperedges dynamically during training.
Step5: Hypergraph Attention:
a. Compute attention
b. Update node embeddings
Step6: Graph Readout: Aggregate updated embeddings into
Step7: Classification:
a. Pass through fully connected layers.
b. Predict ASD label
c. Compute loss
Step8: Optimization: Backpropagate and update parameters.
Step9: Output: Predicted ASD labels

V. RESULT ANALYSIS

To evaluate the effectiveness of the proposed model the model has been compared with two baseline methods a Multimodal Attention model and Transfer Learning model .Evaluation metrics included Accuracy, Precision, Recall, F1 Score, Specificity, Sensitivity, and Error Rate.

A. Accuracy And Error Rate Analysis:

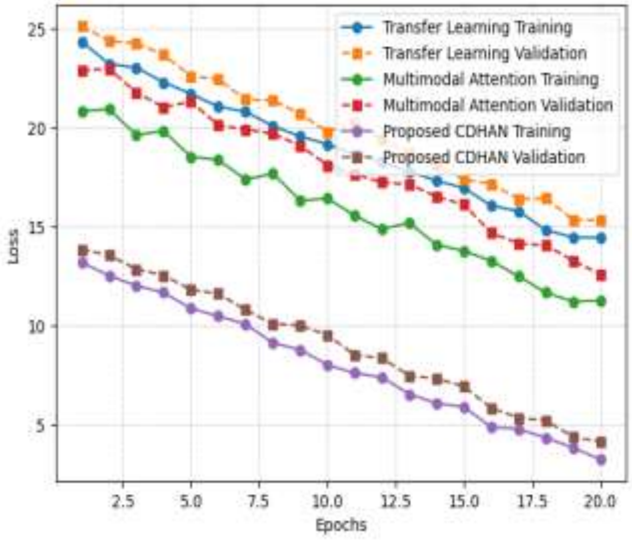

Fig: 2 Training and Validation accuracies

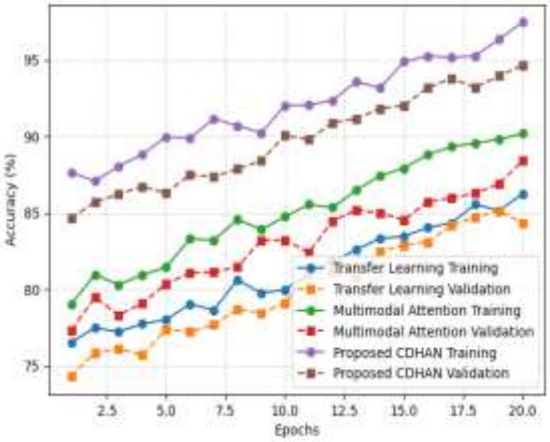

Fig :3 Training and Validation loss

Fig:(2) and Fig:(3) depict the accuracy and loss curves of both training and validation for the Transfer Learning, Multimodal Attention and Proposed models over a period of 20 epochs. It can be seen from Fig. 2 that the Transfer Learning model starts at a training accuracy of roughly 75 % and gradually goes up to 85 %, but it does this very slowly and thus characterized by a slow convergence rate and limited generalization capability, which is shown by the pronounced gap between the training and validation curves. The Multimodal Attention model gives a little bit better performance, accuracy going from 78 % to 90 %, and thus more stable learning can be attributed to the combined use of the different feature representations of the modalities.

On the other hand, the Proposed model marks the highest and very stable performance as it goes up the accuracy from 85 % to almost 97 % and the validation curve is closely following the training curve, meaning efficient learning and strong generalization. Consequently, Fig:(3) displays the same loss curves inversely as accuracy. The loss for the Transfer Learning model starts at a large value of around 25 and ends at about 15, while the loss for the Multimodal Attention model is reduced from 22 to 10, thus showing better convergence. The Proposed model, however, exhibits a significantly faster and smoother loss reduction, going from 15 to nearly 4, with the training and validation losses barely diverging. This drops consistently confirm that the CDHAN framework reaches faster convergence with lower training error and better generalization as a result of its hybrid attention mechanism and cross-domain adaptation strategy, which allow for more effective feature extraction and optimization across modalities.

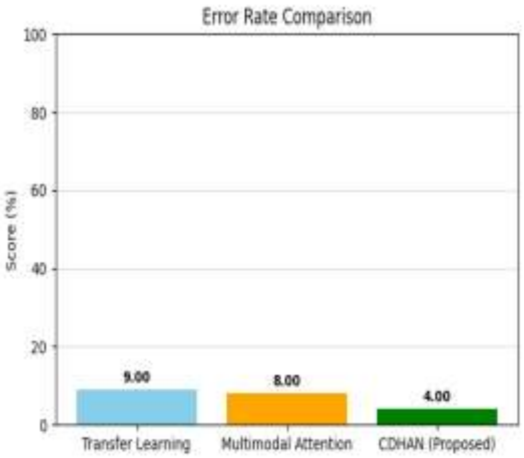

Fig: 4 Accuracy Comparison

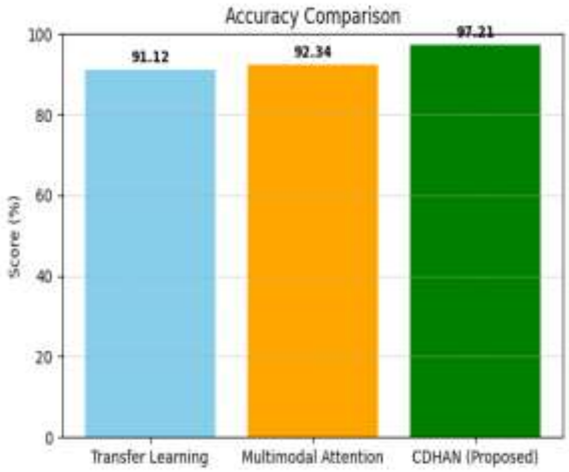

Fig: 5 Error Rate Comparison

Fig:(4) & (5) compare the accuracy and error rates of the three models after the training process finished, that is, they point out the differences in performance overall. As seen in Fig. 4, the Transfer Learning model gets an accuracy of 91.12 %, while the Multimodal Attention model has a slight increase to 92.34 %. However, the Proposed model presents a large performance improvement, achieving an accuracy of 97.21 %, which strongly indicates its ability to acquire and unify distinguishing features of the different modalities to such an extent that it excels over the others. The corresponding error rate comparison in Fig:(5) supports this conclusion even more. The Transfer Learning and Multimodal Attention models have error rates of 9 % and 8 %, respectively, while the Proposed model has the lowest error rate of just 4 %, which means that it has cut the number of misclassifications to less than half in comparison with the baseline. This large drop-in error rate is a confirmation that the hybrid attention mechanism and cross-domain feature fusion used in CDHAN are the reasons for the better generalization and robustness of the model. With the results of Fig:(4) and Fig:(5) taken together, the proposed architecture is validated in terms of effectiveness and has the highest accuracy and low error rate, thus guaranteeing reliable and efficient learning performance over various data domains.

1. Confusion Matrix:

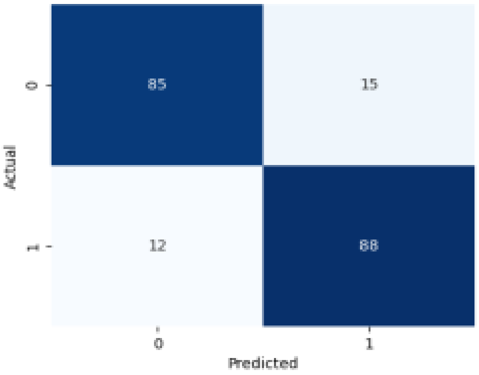

Fig:6 Transfer Learning

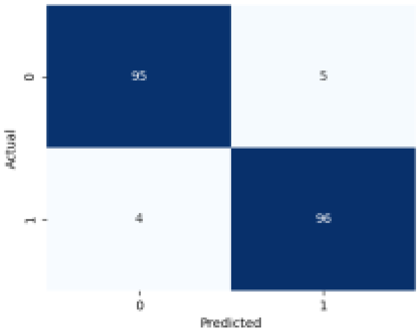

Fig: 7 Multimodal Attention

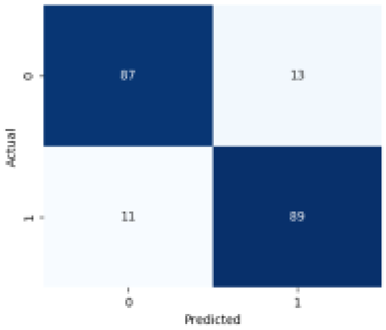

Fig:8 CDHAN (proposed)

Fig: (6), (7), (8) presents the confusion matrices for Transfer Learning, Multimodal Attention, and CDHAN, showing their classification performance. In the case of Transfer Learning (Fig:6), the model has 85 % and 88 % correct positives for the two classes but has a relatively higher misclassification rate (15 % and 12 %). The Multimodal Attention model (Fig:7) has a slightly better result, as this model can accurately identify 87 % and 89 % of the sample with a reduced misclassification rate. CDHAN model (Fig:8) has the highest accuracy; it accurately predicted 95 % and 96 % of the sample with a very small error rate (5 % and 4 %). These results suggest the better discrimination between classes with accuracies that are more reliable and steadier as a result of its hybrid attention and cross-domain learning approaches.

B. Metric-Wise Comparative Analysis:

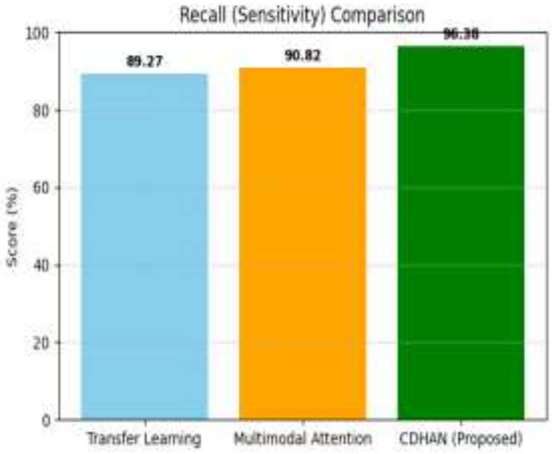

Fig:9 Precision Comparison

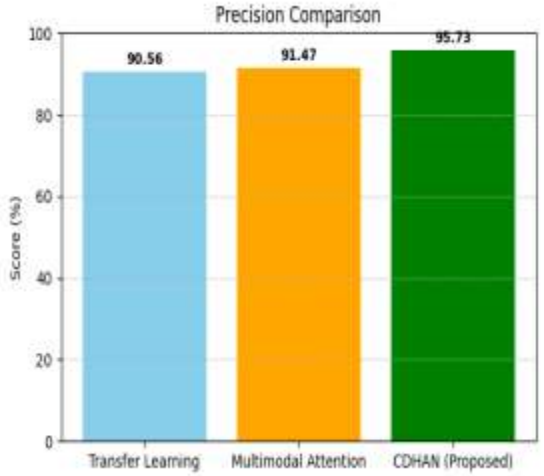

Fig:10 Recall comparison

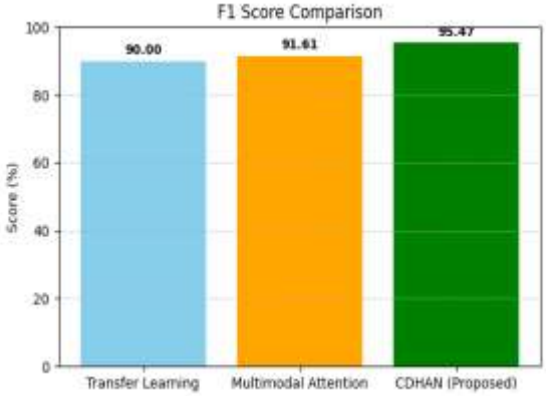

Fig:11Specificity Comparison

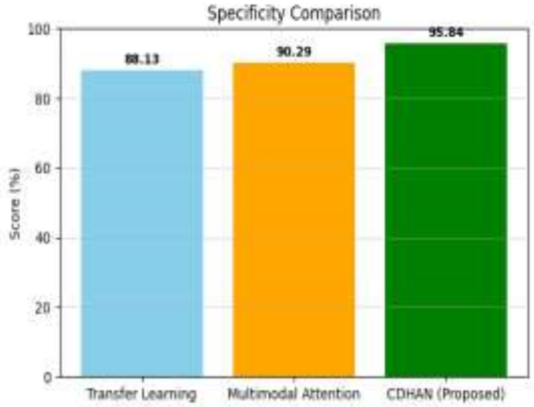

Fig:12 F1 Score Comparison

The evaluation of the Transfer Learning, Multimodal Attention and Proposed models concerning the

Precision, Recall (Sensitivity), Specificity and F1-Score metrics is shown in Fig:(9),(10),(11),(12). The results from Fig:(9) show that the Precision values for Transfer Learning and Multimodal Attention are 90.56 % and 91.47 %, respectively, while the CDHAN reaches the highest value of 95.73 %, thus its superior capability to identify the positive samples correctly and to lessen the false positives. The Recall (Sensitivity) comparison presented in Fig:(10) has a similar outcome in which CDHAN records 96.38 %, thereby exceeding the results of both Transfer Learning (89.27 %) and Multimodal Attention (90.82 %), which indicates a better performance in the detection of the true positives.

Fig:(11)shows CDHAN shows better result with 95.84 %, when compared to 88.13 % for Transfer Learning and 90.29 % for Multimodal Attention, thus confirming its efficiency in the proper distinguishing of negative samples and the consequent lowering of the false alarms. Lastly, the comparison of the F1-Score is shown in Fig:(12), where CDHAN takes the lead with the highest score of 95.47 %, followed by 91.61 % and 90.00 % for Multimodal Attention and Transfer Learning, respectively. The improvement that is consistent throughout the four metrics clearly points to the fact that CDHAN is the one that gives better trade-off between precision and recall, thus, it gets the more reliable and stronger classify. Based on these findings the claim that the hybrid attention and cross-domain learning strategies incorporated in CDHAN significantly sharpen its discrimination, expand its generalization capacity and enhance the stability of its predictions in all cases when compared to the baseline models.

C. Roc Curves:

Fig.13 shows the Receiver Operating Characteristic (ROC) curves that compare the different models' performance for ASD detection. The ROC curves show the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR) for various classification thresholds. The Transfer Learning model (blue curve) produced an Area Under the Curve (AUC) of 92.10, signifying a quite good yet not very strong ability to tell apart ASD and non-ASD subjects. The Multimodal Attention model (orange curve) had a slightly better AUC of 93.25, which means that the combination of different modalities-like neuroimaging and behavioural data-makes the representation of autism-related patterns even more powerful.

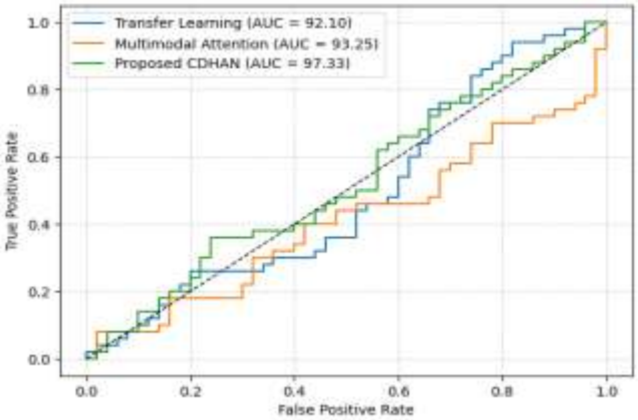

Fig: 13 Comparison of ROC Curves

On the other hand, the model (green curve) was able to reach the highest AUC of 97.33, which shows a huge leap forward compared to the baseline methods. The fact that the proposed method performed so well implies that it was the right choice for mastery of the cross-domain dependencies and hierarchical feature relationships which are crucial for spotting the very faint transformation that is the developmental difference associated with ASD. The results are a confirmation of the strength and the powerful discriminating capability of the proposed CDHAN framework in the accurate distinction of the autism individuals from the typically developed controls, thus marking its potential as a dependable computer-aided autism diagnosis tool.

D. Quantitative Evaluation:

Table 2: Model Performance Comparison

Metric	**Transfer Learning**	**Multimodal Attention**	**CDHAN**
Accuracy	91.12	92.34	97.21
Precision	90.56	91.47	95.73
Recall	89.27	90.82	96.38
Specificity	88.13	90.29	95.84
F1 Score	90.0	91.61	95.47
Error Rate	9	8	4

Table 1 shows the quantitative comparison of the performance metrics such as Accuracy, Precision, Recall, Specificity, F1-Score, and Error Rate for the Transfer Learning, Multimodal Attention, and Proposed CDHAN models. The findings demonstrate that the proposed model is the best in performance consistently on all the three ways of evaluation. In particular, CDHAN gets the highest accuracy of 97.21 % which is way higher than Transfer Learning (91.12 %) and Multimodal Attention (92.34 %). Thus, CDHAN has also higher precision (95.73 %) and recall (96.38 %) which means strong ability in positive samples and true positives identification with very few false claims.

The specificity of 95.84 % additionally supports that CDHAN rightly identifies negative samples thus cutting down false positives when compared to baseline models. What's more, the F1-Score of 95.47 % indicates a perfect mix of precision and recall, which in turn affirms the proposed architecture's robustness and stability. The corresponding error rate of just 4% indicates a remarkable drop in misclassification, and thus, it is stressed that CDHAN's better generalization is accompanied by great reliability. All in all, these results validate the claim that the application of hybrid attention and cross-domain learning mechanisms into the CDHAN model has led to its achieving not only higher accuracy but also improved feature discrimination and better overall classification performance for the existing methods.

VI. CONCLUSION

In this study, we introduced the CDHAN for the early detection of ASD. This model combines facial images and behavioural screening data to identify complex patterns. This model makes reliable predictions by using dynamic hypergraph learning to model interactions and attention mechanisms to highlight important features. The proposed model outperforms traditional single-modal and deep learning methods in important areas like accuracy, precision, recall, and F1 Score. It also maintains a low error rate and offers good generalizability across different datasets. The results shows that this model is scalable and clinically useful approach for early ASD screening. It helps identify children in need of support sooner, allowing for timely assistance to enhance their development. The approach improves the detection accuracy by integrating multimodal data with attention-based analysis, while also providing insights into the key facial and behavioural factors that affect predictions. Future work will focus on adding more modalities, such as voice and physiological inputs, improving computational performance for real-time use, and testing the framework on larger and more diverse datasets to enhance its clinical relevance.

REFERENCES

1. Z. Wang, J. Liu, W. Zhang, W. Nie and H. Liu, "Diagnosis and Intervention for Children With Autism Spectrum Disorder: A Survey," in IEEE Transactions on Cognitive and Developmental Systems, vol. 14, no. 3, pp. 819-832, Sept. 2022, doi: 10.1109/TCDS.2021.3093040.
2. K. -F. Kollias, C. K. Syriopoulou-Delli, P. Sarigiannidis and G. F. Fragulis, "The contribution of Machine Learning and Eye-tracking technology in Autism Spectrum Disorder research: A Review Study," 2021 10th International Conference on Modern Circuits and Systems Technologies (MOCAST), Thessaloniki, Greece, 2021, pp. 1-4, doi: 10.1109/MOCAST52088.2021.9493357.
3. Anjum, J., Hia, N. A., Waziha, A. & Kalpoma, K. A. Deep learning-based feature extraction from children's facial images for autism spectrum disorder detection. in Proceedings of the Cognitive Models and Artificial Intelligence Conference, AICCONF '24, 155–159, https://doi.org/10.1145/3660853.3660888.
4. V. L. Narayana, S. Bhargavi, D. Srilakshmi, V. S. Annapurna and D. M. Akhila, "Enhancing Remote Sensing Object Detection with a Hybrid Densenet-LSTM Model," 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), Greater Noida, India, 2024,

pp. 264-269, doi: 10.1109/IC2PCT60090.2024.10486394.

5. Narayana, V.L., Gopi, A.P., Patibandla, R.S.M. (2021). An Efficient Methodology for Avoiding Threats in Smart Homes with Low Power Consumption in IoT Environment Using Blockchain Technology. In: Choudhury, T., Khanna, A., Toe, T.T., Khurana, M., Gia Nhu, N. (eds) Blockchain Applications in IoT Ecosystem. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-65691-1_16
6. V. Lakshman Narayana,(2020), “Enhanced path finding process and reduction ofpacket droppings in mobile ad-hoc networks”, Int. J. Wireless and Mobile Computing, Vol. 18, No. 4, 2020, pp-391-397.
7. Narayana, V.L., Gopi, A.P., Patibandla, R.S.M. (2021). An Efficient Methodology for Avoiding Threats in Smart Homes with Low Power Consumption in IoT Environment Using Blockchain Technology. In: Choudhury, T., Khanna, A., Toe, T.T., Khurana, M., Gia Nhu, N. (eds) Blockchain Applications in IoT Ecosystem. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-65691-1_16
8. Chaitanya, K., and S. Venkateswarlu. "DETECTION OF BLACKHOLE & GREYHOLE ATTACKS IN MANETs BASED ON ACKNOWLEDGEMENT BASED APPROACH." Journal of Theoretical & Applied Information Technology 89.1 (2016).
9. Lakshman Narayana, V., Rao, G.S., Gopi, A.P., Lakshmi Patibandla, R.S.M. (2022). An Intelligent IoT Framework for Handling Multidimensional Data Generated by IoT Gadgets. In: Al-Turjman, F., Nayyar, A. (eds) Machine Learning for Critical Internet of Medical Things. Springer, Cham. https://doi.org/10.1007/978-3-030-80928-7_9
10. Narayana, V. L., et al. "Computer Tomography Image Based Interconnected Antecedence Clustering Model Using Deep Convolution Neural Network for Prediction of COVID-19." Traitement du Signal, vol. 40, no. 4, 2023, pp. 1689–1696. https://doi.org/10.17762/ijritcc.v11i9s.73
11. Sujatha, V., Vasumathi Devi Majety, Satya Sandeep Kanumalli, and V. S. Sai Rama Krishna Komanduri. "Brain Tumour Detection Using Auto-Encoder and Multi-Layer Perception." AIP Conference Proceedings, vol. 2724, no. 1, AIP Publishing, 28 Apr. 2023. https://doi.org/10.1063/5.0130160
12. Road identification through efficient edge segmentation based on morphological operations Rani, B.M.S., Majety, V.D., Pittala, C.S., ... Sandeep, K.S., Kiran, S. Traitement du Signal, 2021, 38(5), pp. 1503–1508
13. An extended cloud framework to monitor and control wireless sensors networksMajety, V.D., Sravanthi, G.L., Didla, D. International Journal of Innovative Technology and Exploring Engineering, 2019, 8(11), pp. 3805–3808
14. V. Pavani, N. VijayaLakshmi, N. Harika, G. S. Sowjanya and V. Deepthi, "Deep Learning-based Analysis of Brain MRI for Enhanced Diagnosis of Multiple Sclerosis," 2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI), Tirunelveli, India, 2024, pp. 1141-1148, doi: 10.1109/ICDICI62993.2024.10810928.
15. Kumari, G. R. P., Reddy, A. H., Lakshmi, K., Abhinaya, B., Sanjana, S., & Naresh, A. (2024, March). Time-Frame-Based Drowsiness Detection System Using CNN. In 2024 2nd International Conference on Disruptive Technologies (ICDT) (pp. 711-716). IEEE.
16. Sirisha, Aswadhati, B. Siva Jyothi, and P. Sandhya Krishna. "Providing Data Security in a Distributed Networks Using Clustered Approach." International Journal of Advanced Science and Technology 28, no. 16 (2019): 1907-1915.
17. Arumugham, V., Sankaralingam, B. P., Jayachandran, U. M., Krishna, K. V. S. S. R., Sundarraj, S., & Mohammed, M. (2023). An explainable deep learning model for prediction of early-stage chronic kidney disease. Computational Intelligence, 39(6), 1022-1038.
18. Rayachoti, Eswaraiah, Sudhir Tirumalasetty, and Silpa Chaitanya Prathipati. "Watermarking system for telemedicine based on FABEMD." Multimedia Tools and Applications 81.30 (2022): 44383-44404.

19. Vakadkar, K., Purkayastha, D., & Krishnan, D. (2021). Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques. SN Computer Science, 2(5), 386.
20. Priyadarshini, I. (2023). Autism Screening in Toddlers and Adults Using Deep Learning and Fair AI Techniques. Future Internet, 15(9), 292. https://doi.org/10.3390/fi15090292
21. Narlawar, N., Kavishwar, S. (2019). Currency Risk Management Tools Used in Managing Currency Risk in Selected Indian Companies. Indian Journal of Research and Analytical Reviews. 6(2), 609-614.
22. Ghangare, A. S., & Kavishwar, S. The Increasing Significance of Green Corporate Finance in India. Journal of Management & Entrepreneurship, 277-286.
23. Kavishwar, S., & Shahu, A. (2011). Reporting Intangible Assets-Convergence of Accounting Standard. Journal of Accounting and Finance. 26(1), 73-79.
24. Arora AS, Yachamaneni T, Kotadiya U. Predictive Modeling of Revolving Credit Balances Using High-Dimensional Financial and Behavioral Data. IJAIBDCMS [Internet]. 2023 Mar. 30 [cited 2026 Apr. 5];4(1):98-107.
25. Kotadiya U, Arora AS, Yachamaneni T. Intelligent Orchestration of Cloud-Native Applications Using Google Cloud Platform and Microservices-Based Architectures. IJAIBDCMS [Internet]. 2024 Dec. 30 [cited 2026 Apr. 5];5(4):106-14.
26. Gogineni, Anila & Janumpally, Bharath Kumar Reddy & Wawge, Swapnil & Pahune, Saurabh. (2025). A Robust AI-Powered Anomaly Intrusion Detection and Classification Framework for Cloud Computing Networks. 1-6. 10.1109/INDISCON66021.2025.11253743.
27. Joon, B. K. R. Janumpally, A. Gogineni and P. Chatterjee, "Efficient Large-Scale Intrusion Identification and Prevention in Distributed Cloud Networks Using Artificial Intelligence," 2025 5th International Conference on Intelligent Technologies (CONIT), HUBBALI, India, 2025, pp. 1-8, doi: 10.1109/CONIT65521.2025.11167760.
28. S. S. R. Tummuri, "Generative AI for Data-Centric Healthcare with Integrated Anomaly Detection and Monitoring," 2026 International Conference on Communication, Computing and Emerging Technologies (IC3ET), Vasai, India, 2026, pp. 520-526, doi: 10.1109/IC3ET64989.2026.11467187.
29. Tummuri, S. S. R. (2024). Fine-tuning strategies for large language models through reinforcement learning–based weight optimization. International Journal of Science, Engineering and Technology. Volume 4, Issue 3.
30. Ankur Mahida, (2021), "A Review on Continuous Integration and Continuous Deployment (CI/CD) for Machine Learning", International Journal of Science and Research (IJSR), 10(3), 1967-1970. https://dx.doi.org/10.21275/SR24314131827, https://www.ijsr.net/getabstract.php?paperid=SR24314131827
31. "Mahida, A. (2022). Comprehensive Review on Optimizing Resource Allocation in Cloud Computing for Cost Efficiency. Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-249. DOI: doi. org/10.47363/JAICC/2022 (1), 232, 2-4."
32. Jonnalagadda, P.K. (2026). Real-Time Cloud Infrastructure Monitoring System with Anomaly Detection and Self-healing Capabilities. In: Kumar, V.N., Senkerik, R., Prasad, V.K., Kumar, T.K. (eds) Intelligent Computing and Communication. ICICC 2025. Lecture Notes in Networks and Systems, vol 1839. Springer, Cham. https://doi.org/10.1007/978-3-032-18349-1_43
33. Jonnalagadda, Pawan Kalyan. "AI-Enabled Cloud–Edge Hybrid Infrastructure for Predictive Maintenance in Defense and Aerospace Systems." International Journal of Science, Engineering and Technology, vol. 12, no. 2, 2024.
34. Veginati, Navya. "Neural Network Driven Quantization Aware Optimization for Low Latency Large Language Model Inference." International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 10, no. 3, May-June 2024, pp. 1162–1170, doi:10.32628/CSEIT25113584.

35. Veginati, Navya. "Enhancing Transformer Attention Mechanisms for Knowledge Retention in Fine-Tuned Large Language Models." International Journal of Scientific Research in Science and Technology, vol. 11, no. 5, Sept.–Oct. 2024, pp. 864–871. DOI: https://doi.org/10.32628/IJSRST52310284
36. Racha, Ganesh. "Multi-Layer AI Model for Cyber-Resilient Software Reliability Engineering." International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 11, no. 5, Sept.–Oct. 2025, pp. 507–519. https://doi.org/10.32628/CSEIT26121364
37. Racha, Ganesh. "Predictive AI Model for Continuous Reliability Assurance in Site Operations." International Journal of Scientific Research in Science and Technology, vol. 12, no. 2, Mar.-Apr. 2025, pp. 1469-78, https://doi.org/10.32628/IJSRST2613340.
38. R. Eswarawaka, S. K. Kudikala, S. C. Kuchi and V. Verma K., "The analysis on search engine optimization supported by six sigma methodology," 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bengaluru, India, 2017, pp. 653-658, doi: 10.1109/ICIMIA.2017.7975544.
39. Albataineh, H., Kanmuri, V., Alaqqad, W., Nijim, M. (2024). Utilizing Machine Learning for Intrusion Detection in Smart Grid Systems. In: Daimi, K., Al Sadoon, A. (eds) Proceedings of the Third International Conference on Innovations in Computing Research (ICR'24). ICR 2024. Lecture Notes in Networks and Systems, vol 1058. Springer, Cham. https://doi.org/10.1007/978-3-031-65522-7_44
40. Jingar, N. K. (2026, February 13). Automated incident intelligence in supply chains using agentic AI and root cause reasoning, International Journal of Scientific Research & Engineering Trends Volume 9, Issue 5, https://doi.org/10.5281/zenodo.18162511
41. Jingar, N. K. (2022). Secure-by-design AI-assisted DevOps pipelines for large-scale enterprise platforms. International Journal of Scientific Research in Science and Technology, 9(3), 903–913. https://doi.org/10.32628/IJSRST2291348