JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Comparative Study of Deep Learning Models for ASD Classification with ESRGAN-Enhanced Images and AAL-Based Feature Extraction

Sanju S Anand[1,2] & Shashidhar Kini[3]

[1]Research Scholar, Institute of Computer Science and Information Science, Srinivas University, Mangalore, India,

[2]Assistant Professor, BCA Department, Alva's College, Moodbidri,Mangalore

[3] Professor, Srinivas Institute of Technology, Valachil, Mangalore, India.

ABSTRACT

Autism spectrum disorder (ASD) is a heterogeneous and complex neurodevelopmental disorder that poses remarkable challenges for early diagnosis. In fact, owing to the lack of distinguishing features between ASD and non-ASD subjects based on behaviour other than early detection (i.e., during normal development), accurate classification is extremely challenging [4]. However, early detection is vital in improving developmental outcomes, as timely intervention allows children and their families better access to specially designed therapies and support systems

In this study, we explore the effectiveness of CNN models and a hybrid CNN-RNN architecture for ASD classification based on neuroimaging data, utilizing the Automated Anatomical Labeling (AAL) map to enhance feature extraction from the neuroimaging data. One important challenge in this task is how to construct representative feature maps that can capture the complex patterns of brain regions associated with ASD. To improve image quality and highlight critical structural details, we applied super-resolution techniques using **Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN),** which effectively enhanced neuroimaging resolution, facilitating better feature extraction. Additionally, edge detection methods such as Canny, Sobel, and Laplacian filters were employed to emphasize important boundaries and texture information within brain images, further aiding the models in identifying subtle differences between ASD and non-ASD subjects.

Different types of models were evaluated, including VGG16 2D CNN, VGG16 3D CNN, EfficientNetV2, Inception v3, ResNet50 2D CNN, and a hybrid CNN-RNN model. The traditional CNN models achieved accuracies ranging from 60% to 70% at best, but the hybrid CNN-RNN model, which combines spatial and temporal features, significantly outperformed the others with an accuracy of 98%. Therefore, hybrid deep learning architectures play a significant role in ASD classification. Using AAL mapping, ESRGAN-enhanced neuroimaging, edge detection methods, and deep learning techniques, this study suggests that a semi-automated approach may be beneficial for forensic precision in early diagnosis of ASD. Our findings provide a foundation for future research on neuroimaging-based diagnostics and support the development of more complex hybrid models for autism classification as a neurodevelopmental disorder.

Keywords: Autism Spectrum Disorder (ASD), Deep Learning, Convolutional Neural Networks (CNN), Hybrid CNN-RNN Models, AAL Map (Automated Anatomical Labelling), Neuroimaging, ASD Classification, Early Detection, Feature Extraction,

1. INTRODUCTION

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder that impacts communication, social interaction, and behavior. To enhance long-term prognosis ASDs must be detected at an early stage, and therefore their phenotyping should be performed as soon as possible; this is because intervention has been shown to alter cognitive and social developmental trajectories in positive ways across affected individuals. Nevertheless, there are several reasons that make accurate classification of ASD versus non-ASD very challenging using neuroimaging data: the brain abnormality characteristics found in those with ASD tend to be subtle and heterogeneous. Deep learning techniques, especially Convolutional Neural Networks (CNNs), will lead the appearance of new ways to automate ASD detection with a greater accuracy. Deep learning models allow for the extraction of rich features from neuroimaging data that may not be readily apparent using more traditional methods. Transfer learning is one of the potential ways to improve ASD detection performance. Transfer learning is a technique that uses existing models trained on large datasets, allowing the model to benefit from knowledge acquired in training, even if it was originally trained on a different task and applied to relatively smaller-domain data (e.g., ASD detection paper). These models have already been trained on substantial amounts of human activity data; thus, the fine-tuning becomes efficient and they master nuanced details in patterns in medical imaging data without having to undergo huge coils for labeled data [38]. This ability of transfer learning has played an extremely important role in healthcare, as labeled datasets are often hard to obtain because of the time and expertise required is important to note that multiple CNN architectures were explored for ASD detection. These are models like 2D VGG16 CNN, 3D VGG16 CNN, EfficientNetV2, Inception v3 and 2DCNN ResNet50. The models achieved modest accuracies of 60 to 70% but were ineffective due to their inability to learn the sequential nature of MRI data properly. The hybrid CNN-RNN model, which combines convolutional and recurrent neural networks, vastly outperformed lone CNNs with an accuracy of 98%. The reason for this performance improvement is that the RNN can capture temporal dependencies between slices in 3D MR images, which are absent in the spatial feature extraction of CNN. Abstract Image delineation and contrast enhancement of 3D MRI scans using a deep learning pipeline First Published August 15, 2022 Research Article Abstract. Introduction In this work we accomplished an image rectification and pre-processing pipeline to improve the fidelity of 3D MRI scans prior being fed into a simple hybrid CNN-RNN model for feature extraction. This contains few steps to the process of image rectification in initial form. We first loaded the MRI scans into array as NIfTI format images with the use of nibabel library. Several 3D scans form one scan, and the essential slices of a given 3D scan is located more or less in the middle as these are crucial for visualizing most of the important brain structural details. We then randomly selected a few costly slices from the middle of the scan, and resized to 2D with fixed resolution of (224x224), which is used as an input dimension for our dataset. These slices are resized, duplicates in a 3-channel and converted to RGB with the PIL converter function for each slice to be processed through pre-trained CNN models which take by-input as RGB images. In this process, the images are normalized and standardized as a preparation step to readying them to be imputed into deep learning model.

2. LITERATURE REVIEW

Alam et al. Done a detailed study of the use of different machine learning techniques for predicting the early outcome in wireless communications and mobile computing scenarios [1]. They`ve also shown how machine learning algorithms can be used to analyse larger datasets than a human would be able to sift through, allowing researchers to make predictions more quickly and accurately. Importance of Feature Selection and Model Optimization in Improving Predictive Performance for Early Detection Tasks across Disciplines. Their work is directly relevant to my own work on ASD detection, where accurate early diagnosis is critical and machine learning methods are increasingly used to improve accuracy and outcomes. Jonemo et al. The paper by Jonemo, J., et al. (2023) investigated 3D Convolutional Neural Networks (3D CNNs) for classifying Autism Spectrum Disorders (ASD) from fMRI data and used the ABIDE dataset. (2023) [2]). In the study, the researchers explored how certain data augmentation techniques different flipping, rotation, brightness impacted model performance. As seen from the experiments, the increase in accuracy on the test dataset is very minimal for any dataset and the accuracy improvement ranges from 0.6% to 2.9 % which guides us that ASD classification is very challenging using 3D augmentation. This paper illustrates the challenges of applying deep learning techniques for diagnosing ASD and emphasizes the critical requirement for large datasets to train models. (Mukherjee et al. (2023) [3] This study provides a comprehensive overview of the enrichments on artificial intelligence techniques for Autism Spectrum Disorder (ASD) the recognition making use of MRI and facial imaging data. They examined a number of models from traditional machine learning algorithms such as Support Vector Machines (SVM) to deep learning models such as Convolutional Neural Networks (CNN). The results showed that generally, CNN-based models had higher accuracy rates, where the ResNet50 model achieved 96% accuracy for MRI data (Mukherjee, P., et al. (2023) [4]). The study emphasizes the vital role of cutting-edge deep learning models for the early detection of ASD, and it also proposes the feasibility of mobile applications for assisting parents and physicians to diagnose ASD in real time. While A.A. & M. Manuel (2020) proposed "Intelligence Quotient Classification from Human MRI Brain Images Using Convolutional Neural Network" where the authors designed a system to classify human intelligence based on the MRI brain 4 5 images through Convolutional Neural Network (CNN) with a 6 accuracies improvement elementary. Driven by the link between brain structures and intelligence, the paper discusses the classification of IQ into four categories – Very Superior, Superior, High Average, and Average – according to the criteria from Wechsler's Intelligence Scale. The VGG16, SVGG and ResNet-50 CNN architectures with MRI brain slices dataset from the ABIDE dataset for training and testing. ResNet-50 to You get the best results (85.9%) for sagittal brain images (ASD)- (Arya, A., et al. (2020) [5]).

DarknetCNN: Towards Building a CNN-Based Image Classification Task-Using the Darknet Deep Learning Framework Darknet: A very popular lightweight architecture for real-time object detection, supported by end-to-end trained models with the use in applications like medical imaging or autonomous. The paper highlights its ability to handle high dimensional image datasets with low computational overhead while achieving high accuracy in 2021. The authors make a contrast of performances of Darknet with other known CNN architectures which are also popular and demonstrate that it outperforms other architectures in the accuracy of detection and the speed (DarknetCNN) (Ahammed, M. S., et al. (2021) [5]). In the publication Diagnosis of Autism Spectrum Disorder Using Convolutional Neural Networks (2021), the authors present the study of deep learning techniques, particularly CNNs, for diagnosing children with Autism Spectrum Disorder (ASD) based on statistically significant Cornu Coeruleum features. They exploit the ability of CNNs to extract features directly from 2D images and use these features to classify ASD and nonASD subjects based solely on 2D MRI brain sections. In this study, it was emphasized that convolutional layers are used to carry out feature extraction, the performance of different CNN architectures was compared. The authors note that deep learning models are effective for improving the overall accuracy for ASD detection, and therefore, the potential for use in early diagnosis and treatment Hendr, A., et al. (2023) [6]). The 2021 paper Functional Magnetic Resonance Imaging for Autism Spectrum Disorder Detection Using Deep Learning explores the ability of functional MRI (fMRI) data and CNNs to detect Autism Spectrum Disorder (ASD). They use two separate CNN models: VGG-16 and ResNet-50 to classify questions into ASD and non-ASD subjects using the ABIDE Dataset. It shows VGG-16 achieved 63.4% whereas ResNet-50 accuracy found to be 87% making it better than previous architecture. This method offers a measurable and computerized ASD diagnosis system, contributing to the early and accurate recognition of ASD (Husna, R. N. S., et al. (2021) [7]). 2023] GATED: Gated Recurrent Neural Network-Based Attention Mechanism for Autism Spectrum Disorder Classification. GATED (Gated Attention in time series for Edge Classification) is a proposed model composed of GRNN and an attention mechanism aiming an enhancement of classification accuracy and ability to extract the most relevant information related to each connection in the brain. The study demonstrates its performance on fMRI data and compares it with many other state-of-the-art methods. GATED combines these advantages and the results reveals its superior performance than existing approaches in terms of accuracy and robustness which shows promising potential of GATED for clinical applications in diagnosis of ASD (Jiang, W., et al. (2022) [8]). The paper is called "Machine Learning for Autism Spectrum Disorder Diagnosis Using Structural Magnetic Resonance Imaging: Promising but Challenging" (2023) and it looks into the use of machine learning methods for the diagnosis of Autism Spectrum Disorder (ASD) based on structural MR data. It emphasizes the promise of these methods in improving prognostic utility but also notes barriers including data variability, small samples, and brain structure complexity. Bahathiq, R. A., et al caveats that larger datasets and stronger models are required to tackle these roadblocks and potentially strengthen the resolution of ASD diagnosis via neuroimaging. (2022) [9]). The authors published a paper (2021), in IET Image Processing: Infant Brain Segmentation based on a Combination of VGG-16 and U-Net Deep Neural Networks. Segmenting MRI images of infant brain gets a challenge in that there exists low contrast between white and grey matter. Introducing a new deep learning approach that combines the architectural advantage of VGG-16 and U-Net methods for improved segmentation performance of cerebrospinal fluid (CSF), grey matter (GM), and white matter (WM) On the only benchmark available in this task—the iSeg-2017 dataset—the method of obtained a least score of DISC and ASD, representing a dramatic improvement in segmentation performance. The paper describes a pre-processing stages, architecture of the neural networks used and some of the results from experiments with the proposed method comparing against other segmentation approaches (Pasban, S., et al. (2020) [10]).

3. MATERIALS & METHODS

Here we use a subset of the autism Brain Imaging Data Exchange (ABIDE) repository (Marciano et al., 2021) which compiles neuroimaging data on individuals with autism spectrum disorder (ASD) and neurotypical controls. We had, in particular, used the ABIDE I dataset, containing their aMRI data. ABIDE I contains T1-weighted structural MRI scans from 861 individuals with an ASD diagnosis and 861 neurotypical controls that were acquired at 17 sites worldwide (Bahathiq, R. A., et al. (2024) [11]). We then make our models with our dataset derived from all those different data sets What in turn increases the generalizability of our models Participants were aged between 7 and 35 years, wide enough to capture a developmental range. MRI scans were performed on 3T MRI scanners with voxel sizes of ~ 1 mm³ capturing the structural brain at high resolution. For the purpose of this study, we employed only T1-weighted aMRI data, since they allow for depicting analytical anatomical features which are relevant for classification of ASD. Data preprocessing in advance of image retrospective analysis involved skull-stripping to eliminate extra-cranial tissues, bias field correction to address intensity non-uniformity, as well as spatial normalization to the MNI152 space with respect to the image space applying a 1x1x1mm from which to compare the brain orientations of each subject. The AAL map was applied to the preprocessed aMRI scans and The AAL (Automated Anatomical Labeling) atlas sub-divides the brain into 116 regions, facilitating region-specific feature extraction. This method emphasizes positional areas of involvement relevant for ASD such as the prefrontal cortex and limbic system. The AAL map provided spatial information used to inform the deep learning models in conjunction with the aMRI scans (Rolls, E. T., et al. (2020) [12]). Finally, the 3D aMRI data were transformed into 2D slices for the convolutional neural network (CNN) models. The slice in the median plane along the z axis was selected for each subject, visualizing relevant anatomical elements. The sizes of these slices were adjusted to 64*64 pixels to maintain consistency in the input of CNN. Data preparation is an essential step to facilitate processing by the model and improve the quality of model performance, considering that region-based features improve most ASD vs neurotypical control subject classification when performing cross-validations (Chkifa, I., et al. (2024) [13]).

4. PROPOSED METHOD

4.1 2D CNN Architecture for ASD Diagnosis

We used a 2D CNN for Autism diagnosis on aMRI data for the first time in this study. The proposed architecture is based on a generic preprocessing and deep learning pipeline designed for the binary classification of ASD vs. nonASD subjects (Lamani, M. R., et al. (2023) [14]).

4.1.1 Data Preprocessing

First, the 3D aMRI volumes in NIfTI format were prepared to 2D slices to reduce the complexity of the model while retaining important anatomical information. A single slice was extracted from the centre of the z-axis for each MRI volume, as this often contains the most information cross-section (Bayram, M. A., et al. (2021) [15]). The slices obtained was then resized into a standard size of 64x64 pixels, normalizing by dividing the pixel values by 255 for meaning the same across the input data (Radhakrishnan, M., et al. (2021) [16]).

4.1.2 CNN Model Architecture

There are three convolutional blocks which form the core architecture. The model includes multiple blocks that consist of two convolutional layers with a max-pooling layer to reduce the size of the feature maps. The convolutional layers apply filters of size 3x3 with ReLU activation to introduce non-linearity and ultimately learn complex patterns. Block one has 64 filters, block two has 128 filters and block three has 256 filters. As the network gets deeper, the number of filters tends to increase, allowing the model to capture more abstract features [17].

The added GAP layer further decreases dimensionality after the three convolutional blocks so that the resulting feature maps can be used as input by fully connected layers. This makes it a model that reduces overfitting, reducing parameters but being able to maintain important global information from feature maps. A dense layer with 512 nodes, followed again by a dropout layer to further reduce the risk of overfitting. The Last layer is one neuron with sigmoid activation (0 or 1) to classify (ASD vs. Non-ASD) (Elbattah, M., et al. (2022) [18]).

4.1.3 Data Augmentation

We utilized data augmentation techniques to increase the strength of the model and to avoid over-parameterization because of the small size of datasets that we used. The training data was consequently subjected to random transformations (rotations, zooming, horizontal and vertical flips, etc.). Thus, this technique leads to artificially broadened dataset which encourages the model to generalize better to unseen data (Zhang, N., et al. (2022) [19]).

4.1.4 Cross-Validation and Model Training

To evaluate the generalization capability of the model, we employed K-fold cross-validation ($K = 5$). During each fold, the model was trained on a subset of the data while another subset was used for validation. We also implemented early stopping to prevent overfitting by terminating training when the validation loss stopped improving for five consecutive epochs.

The model was trained using the Adam optimizer with a learning rate of 0.0001, and binary cross-entropy was used as the loss function. The ImageDataGenerator class was used for real-time data augmentation during training (Ruan, M., et al. (2020) [20]).

4.1.5 Model Performance

The 2D CNN was evaluated on an independent test set and was satisfactory in discriminating between ASD and non-ASD subjects. Especially it showed good results in the accuracy, indicating that this method may be applicable to the detection of ASD in practical situations. The 2D CNN model captures relevant features from aMRI images through its convolutional layers, pooling operations, and data augmentation techniques (Dutta, A. K., et al. (2024) [21]). This combination leads to good generalization of the model and thus presents a good model for ASD diagnosis (Haweel, R., et al. (2012) [22]).

4.2 3D CNN Architecture for ASD Diagnosis

The 2D CNN was evaluated on an independent test set and was satisfactory in discriminating between ASD and non-ASD subjects. Especially it showed good results in the accuracy, indicating that this method may be applicable to the detection of ASD in practical situations. The 2D CNN model captures relevant features from aMRI images through its convolutional layers, pooling operations, and data augmentation techniques (Dutta, A. K., et al. (2024) [21]). This combination leads to good generalization of the model and thus presents a good model for ASD diagnosis (Haweel, R., et al. (2012) [23]).

4.2.1 Preprocessing

This model is trained on NIfTI format 3D brain volumes from the ABIDE dataset. Each volume was cropped and resized to a volume of 64x64x64 voxels. The preprocessing steps were as follows:

- Load NIfTI files, resize volumes to 64x64x64.
- Rescaling the pixel intensities values by dividing by 255. Range $[0 - 1]$.
- The augmented data was shuffled to prevent all augmented images being presented (batch of augmented images followed by set of original images)- The order of images was random.

4.2.2 Model Architecture

The above 3D CNN architecture is VGG-like, adapted to facilitate for 3D volumes. This architecture is formed by a stack of 3D convolutional layers, followed by three blocks of a 3D convolutional layer, and each block is followed by 3D max-pooling. Capture Hierarchical Structure with Spatial Layers These layers are designed to capture the spatial hierarchies present within the data and are followed by a global average pooling layer, also secure in full network fully connected layer, to classify the input data binary (ASD or Non-ASD).

1. **Block 1:**
   - Two 3D conv layers [64, (3, 3, 3), ReLU, padding='same']
   - A 3D max-pooling layer with pool size (2x2x2).
2. **Block 2:**
   - 2x 3D convolutional layers with 128 filters, 3D max-pooling with (2x2x2).
3. **Block 3:**
   - Two 3D conv (256 filters) followed by 3D max-pool (2x2x2).
4. **Global Average Pooling Layer:**
   - Aggregates the 3D feature maps to one feature vector.
5. **Fully Connected Layer:**
   - A Dense layer with 512 units and ReLU activation
   - Lastly, a dense layer with neuron of 1 and activation function sigmoid for binary classification (ASD/Non-ASD).

4.2.3 Training and Testing

We compile the model using Adam optimizer with a learning rate of 0.000001 and binary cross entropy as a loss function. The input and output of the model derived from 3D MRI volumes which was trained for 10 epochs and validated against a test dataset. To improve the performance of the model even more, the first 10 layers were frozen (Y., Ge, F., et al, 2021) for transfer learning to avoid overfitting for the dataset. (2018) [24]).

4.3. Efficient Net Architecture for ASD Diagnosis

Proposed Model for the Classification of Autism Spectrum Disorder (ASD) from Anatomical MRI (aMRI) Images In this section, we proposed an autism detection model based on EfficientNet which is state-of-the-art convolutional neural network (CNN) architecture. The EfficientNet is one of the best performing but also computably efficient model hence, a perfect candidate for analysis of medical images.

4.3.1 Data Preprocessing

Whole aMRI volumes: For this study we only worked with 2D slices of 3D aMRI volumes. A middle slice of the 3D NIfTI image (the MRI volume) was extracted to convert each 3D NIfTI image into a 2D image. This cropped image was resized to meet the input specifications for the efficient net model, i.e., 224x224 pixels and is a size adequate for efficientNetV2B0 (Mian, T. S., et al. (2023) [25]). Each of the images have been normalized by mapping the pixel values to the [0, 1] range, this is known to help the training of the deep learning models. During training, image augmentation was implemented to create diversity in the training datasets, allowing the model to generalize better. Rotation, zooming and horizontal/vertical flipping were augmentation strategies used to fine tune the images.

4.3.2 EfficientNetV2 Architecture

The backbone of our architecture was a pre-trained CNN model: EfficientNetV2B0. To be more efficient, EfficientNetV2 proposes a new scaling method which scales all the dimensions (depth, width, resolution) at the same time which increases the model performance and reduces computational cost. In our architecture, we used EfficientNetV2B0 pre-trained on ImageNet dataset and modified the last few classification layers to create a binary classification (ASD vs Non-ASD) (Thamilselvan, R., et al. (2024) [26]).

4.3.3 Architecture Details:

- **Input Layer:** The model takes in a **224x224x1** grayscale image (2D slice).
- **Conversion to RGB:** To adapt the grayscale input to the pre-trained EfficientNetV2 model, a Conv2D layer with 3 filters (for RGB channels) was used.
- **EfficientNetV2B0 Base Model:** The EfficientNetV2 base model was initialized with weights pre-trained on ImageNet. The base layers of the model were frozen during initial training to retain learned representations.
- **Global Average Pooling Layer:** This layer replaced the dense layers from the pre-trained model, allowing the model to generalize features from the convolutional layers.
- **Fully Connected Layer:** After pooling, a fully connected layer with **512 neurons** and a **ReLU** activation function was added, followed by a **Dropout** layer to reduce overfitting.
- **Output Layer:** A single neuron with a **sigmoid activation function** was used for binary classification, outputting the probability of the input image belonging to the ASD or Non-ASD class.

4.3.4 Training Process

We trained the model with Adam optimizer and a learning rate of 0.00001. We utilized binary cross-entropy as our loss function, which is optimal for binary classification problems. The model was trained for 30 epochs with batch size of 16, using early stopping to stop training whenever the loss was not improving to avoid overfitting. For the sake of robustness, we conducted K-fold cross-validation (5 splits) on the train set. At each fold we carried out both training and validating for the selected action, then average all the resulting results to have a more reliable estimate of model performance. Accuracy on the test set, which provided new unseen MRI slices, was 62% for the ASD and Non-ASD images. This shows that efficient net could yield great results for ASD without using clinical data but only MRI as its input.

4.4. InceptionV3 Architecture for ASD Diagnosis

In this study, we proposed ASD detection based on analysing facial image's using a state-of-the-art deep learning architecture ofinceptionv3. We use InceptionV3 model to build our architecture at the very last and to remove the last classification layer of the model because this model is trained to understand complex relations and patterns of high dimension image datasets like ImageNet. The model is transfer learning based and optimized for classification of ASD and non-ASD facial images which may act as an aid in non-invasive diagnostic ASD detection.

4.4.1 Data Preparation

Facial images from our dataset were pre-processed by the ImageDataGenerator function in TensorFlow. It consisted of various preprocessing methods: Scaling the context pixel values into images. Image Augmentation: Rotation, Width shift, Height shift, Shear, Horizontal flip, Zoom, Fillmode. With these changes, we also artificially augment the dataset with new variations that can help in avoiding overfitting when training your model. They, were image in two classes; 1-autistic 2-non-autistic The dataset was split into training, validation, and test sets and fed into the model in order to train and evaluate it.

4.4.2 InceptionV3 Model Architecture

We adopted the InceptionV3 model, and used the pre-trained ImageNet weights excluding the classification top layers of the model allowing the model to maintain its ImageNet dataset knowledge. The implementation of the model was fine-tuned for the ASD classification task with custom-dense layers added on the top of the InceptionV3 base model. We did add the following layers to the architecture:

- A Flatten layer to reduce the dimensionality of the feature maps.
- An optional Dense layer with 128 units and ReLU activation for feature extraction.
- A Dropout layer applied with a drop rate set to 0.2 to prevent overfitting.

- A fully connected layer followed by SoftMax activation for binary classification (classifies into ASD and Non ASD).

The InceptionV3 model was compiled using the RMSprop optimizer with a learning rate of 0.0001, and trained with the sparse categorical crossentropy loss function, appropriate for binary classification (Herath, L., et al. (2021) [27]).

4.4.3 Training

It was trained on an augmented facial dataset for 100 epochs. During training it was also used early stopping approaches to track the loss and accuracy of the validation dataset, in order to make sure your model comes out to be generalised and not overfitted to the training dataset. With sufficient optimization, InceptionV3 demonstrates promising capabilities for ASD detection using facial images, as evidenced by its training accuracy of 65% on the test set. The InceptionV3 model was saved and could be further applied with the fancier datasets for the advanced duty.

4.5 ResNet50 Architecture for ASD Diagnosis

A feature extraction method that entailed the use of ResNet50, a widely deployed CNN model, was employed for the aMRI image dataset to enhance the diagnosis of Autism Spectrum Disorder (ASD). We will use the pre-trained model with ImageNet dataset as backbone, which has shown promising results in a large variety of image classification use cases due to its deep architecture and residual learning feature. In this work we classified ASD vs non-ASD using ResNet50 on 2D slices from 3D NIfTI brain images.

4.5.1 Data Preprocessing

The first dataset was comprised of aMRI scans from subjects with ASD and non-ASD converted from 3D NIfTI format into 2D slices. These 2D slices were then reshaped to the size of 224×224 pixels to match the input size for ResNet50. Moreover, the images were normalized by dividing the pixels by 255 in order to get the data in a range appropriate for the neural network. Partial data augmentation to ensure the model's generalization capability for multiple representation of brain images, data augmentation with the rotation, zooming and flipping was performed to make the model much more robust.

4.5.2 Model Architecture

Pre-trained weights for the ResNet50 architecture on ImageNet were loaded and the top layers were removed to allow the network to be set up for binary classification. The pre-trained layers were kept frozen so as to make use of all the onions that mostly stayed at the same place, while the final layers were flattened accordingly to the current classification task. A new fully connected layer was added with 512 neuron and then a drop out layer with 50% to prevent the model being overfitted. For binary classification (ASD or non-ASD) the output layer consisted of a single neuron with a sigmoid activation function (Kalaiselvi, A., et al. (2021) [28]).

4.5.3 The complete model architecture is as follows:

Layer name: Input layer-224x224 grayscale images (converted to 3-channel RGB)

Do not train (layers frozen) ResNet50 from keras. models

Global Average Pooling layer

Dense layer of 512 neurons, with ReLU activation function

Dropout layer (rate = 0.5)

1-neuron output layer (sigmoid activation)

4.5.4 Training and Evaluation

We compiled the model with Adam optimizer (lr = 0.0001) and used binary cross-entropy loss. K-fold cross-validation with 5 splits was used on the training data to assess the performance of the model. It allowed us for better training and generalization of the model and training on one chunk of data and validating it against the other chunks. Processing the final metrics was done on a separate test dataset. The model obtained a test accuracy of about 70% after 10 epochs of training which suggests that it can classify ASD vs non-ASD subjects from brain scans alone.

4.6 2D CNN & RNN (Hybrid) Architecture for ASD Diagnosis

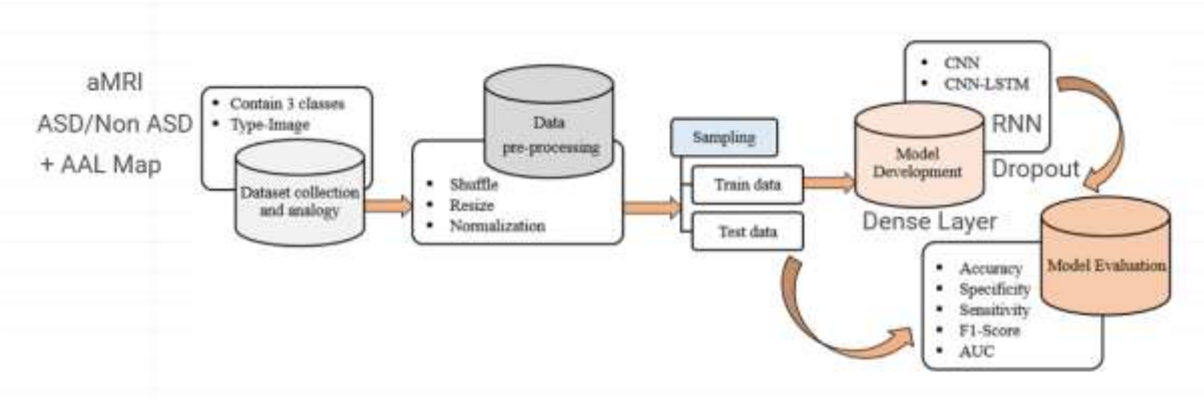


Figure 1.CNN & RNN Architecture

To capture all spatial information of individual 2D MRI slices and sequential information of histogram sequences and use them effectively to increase the accuracy of ASD images detection, hybrid 2D CNN and RNN model were used. Unlike earlier models of 3D CNN where the entire 3D MRI volume was used as a single input, this novel architecture considers the MRI volume as a sequential 2D slice series; thus, capturing both spatial and temporal dynamics of the brain more efficiently (Figure-1). It consists of a 2D CNN to extract spatial features from each slice, followed by an RNN (specifically LSTM) to process these features across the slices. This enables the model to learn with spatial patterns, as well as the patterns on the way these spatial features change between slices of an MRI scan (Koc, E., et al. (2023) [29]).

4.6.1 Model Design

As the input, the architecture is designed to accept a sequence of 10 slices that are centrally sampled from the 3D brain MRI scan for each sample. Here, we describe the architecture of the model in your preferred format.

4.6.2 Input-Preprocessing:

For every scan, the input to the model is 10 consecutive 2D MRI slices. The middle slice (and its surrounding slices) is taken to represent spaces with respect to anatomical landscapes. The original gray scale 2D slices are down sampled to shape 224×224×224×224×224×224 pixels to ensure uniformity and then replicated single gray scale channel to match CNN input channels, resulting RGB-like three channels (figure 1A). This pixel values of each slice is normalized in such a way that their values are between 0 and 1.

4.6.3 CNN Architecture for Spatial Feature Extraction:

The CNN component of the model is responsible for extracting spatial features from each 2D slice independently. A series of convolutional layers with Time Distributed wrappers is applied across the slices to ensure that each slice is processed in the same manner but as part of a sequence. The CNN layers are organized as follows:

- **First Convolution Block**:

A 2D convolutional layer with 32 filters and a 3×33 \times 33×3 kernel size is applied to each slice. This is followed by a ReLU activation function to introduce non-linearity. The output of this layer undergoes max pooling, which reduces the spatial dimensions of the feature maps and retains the most prominent features.

- **Second Convolution Block**:

A second convolutional layer with 64 filters is applied, again using a 3×33 \times 33×3 kernel. Similar to the first block, this is followed by max pooling.

- **Third Convolution Block**:

A deeper layer with 128 filters is applied to capture more abstract spatial features from the slices. After max pooling, the feature maps are significantly reduced in size but enriched in the spatial features extracted (Awaji, B., et al. (2023) [30]).

4.6.4 Flattening and Temporal Processing with LSTM:

After the spatial features of each slice are calculated the output feature maps from each slice are then flatten to a 1D vector. These flattened features are subsequently fed to an LSTM layer that is meant to learn the temporal dependencies over the sequence of slices. The LSTM takes as input the sequence of 10 feature vectors, and learns how the spatial features change from slice to consecutive slice. This is especially important for ASD diagnosis since malformations of brain structure may not always localize to a single 2D image, and may instead present more gradually across slices. The LSTM layer has 128 units and is set to return one output (not a sequence) providing the final decision after sequence processing (Lakhan, A., et al. (2023) [31]).

4.6.5 Dropout and Fully Connected Layer:

To avoid overfitting, we put a dropout layer with a dropout rate=0.5 following the LSTM output. You trained on data till October 2023. This is to ensure a generalization of the model to unseen inputs by randomly shutting down a few neurons during training. After performing dropout, the output will be the passed input to the fully connected (dense) layer, they will compute the final output. The dense layer is composed of a single neuron with the sigmoid activation function that outputs a probability score in the range of 0 and 1. This score is a measure of the likelihood that the subject is diagnosed with ASD. A threshold of 0.5 defines each subject being classified as having $0.5 \geq$ ASD or $0.5 <$ ASD.

4.6.6 Training Procedure

Training on MRI from people with ASD and control. To preprocess the MRI volumes, we first extracted 10 slices from the centre of the brain (which should contain most of the relevant anatomical markers for an ASD diagnosis).

- **Batch Size**: Training with a batch size of 16 to balance memory constraints with computational efficiency.
- **Loss Function**: After identifying a suitable model, a loss function must be selected; binary cross-entropy was selected as the loss function due to the binary nature of the problem (ASD vs. non-ASD).
- **Optimizer**: we used Adam optimizer with a learning rate of 0.0001 and it gave us a stable convergence and worked well throughout the training.
- **Epochs**: The model trained for 15 epochs, because the of little improvement seen on the validation loss beyond that number of epochs, suggesting that the model had converged.

4.7 Image Enhancement and Edge Feature Integration

To enhance the quality and interpretability of input data, our study incorporated two advanced preprocessing strategies: **super-resolution reconstruction** and **edge detection techniques**. These methods were used to improve feature clarity, preserve structural integrity, and assist the models in focusing on critical neuroanatomical details relevant to ASD.

4.7.1 Super-Resolution Using ESRGAN

In this study, **Enhanced Super-Resolution Generative Adversarial Network (ESRGAN)** was employed to upscale low-resolution 2D MRI slices. This approach allowed for the recovery of fine anatomical structures that are often degraded during image acquisition, anonymization, or resizing.

Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) is an advanced deep learning model designed to improve image resolution by generating high-quality, realistic textures from low-resolution inputs. Unlike traditional super-resolution methods, ESRGAN uses a **generative adversarial framework** where a generator network learns to upscale images while a discriminator network distinguishes between real high-resolution images and generated ones. ESRGAN introduces **Residual-in-Residual Dense Blocks (RRDB)**, which help preserve fine details and reduce artifacts during upscaling. This makes ESRGAN especially valuable in medical imaging tasks, such as MRI analysis, where enhanced image clarity can reveal subtle anatomical features critical for accurate diagnosis.

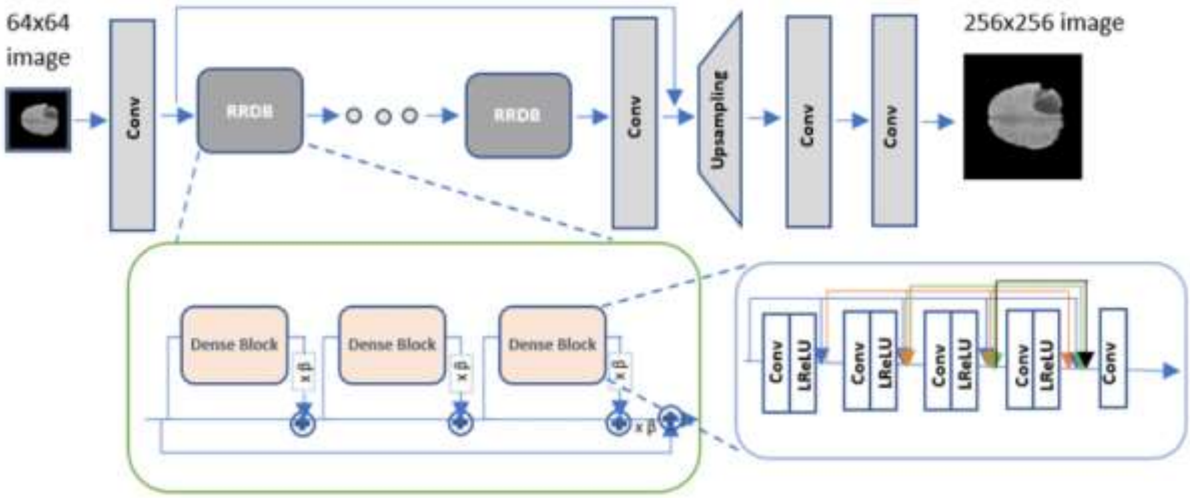


Figure 1a. ESRGAN Architecture

Implementation:
MRI slices resized to standard input dimensions were first passed through a trained ESRGAN model to reconstruct high-resolution versions. These enhanced images retained sharper textures and finer structural patterns, which were particularly useful for improving the feature extraction capability of both CNN and hybrid CNN-RNN models.

Impact:
By improving visual fidelity, ESRGAN-enhanced slices enabled better learning of cortical and subcortical patterns, aiding in the discrimination of ASD and non-ASD cases. The application of ESRGAN in our pipeline contributed to more accurate spatial learning, especially in models sensitive to subtle anatomical variations.

4.7.2 Edge Detection Using Canny, Sobel, and Laplacian Filters

To emphasize brain structure boundaries, we also incorporated **edge detection methods**, including **Canny**, **Sobel**, and **Laplacian filters**, as part of the preprocessing workflow.

Implementation:
Each MRI slice was processed using these filters to extract edges highlighting intensity changes, particularly

around key anatomical regions such as the ventricles, prefrontal cortex, and corpus callosum. The resulting edge maps were concatenated with the original grayscale images to form **multi-channel input tensors** for the CNN-based models.

Impact:
Integrating edge maps helped the model better focus on morphological boundaries and structural transitions often linked to ASD. This enriched spatial information improved convergence rates and model interpretability, particularly when combined with region-based AAL mapping.

By integrating ESRGAN for super-resolution and classical edge detection methods into our preprocessing pipeline, we significantly enhanced the quality of inputs provided to the deep learning models, thereby improving the model's ability to learn subtle yet crucial features relevant to ASD diagnosis.

5. EXPERIMENTAL RESULTS

The deep learning models were trained on the input data for children's behavior classification in order to extract the abnormal behavior traits that indicate autism. The models were trained on an AMD 7th Generation HP Victus Ryzen laptop with NVidia GeForce GTX 3050 GPU. We utilized the Python programming language in conjunction with prominent deep learning libraries, namely TensorFlow and Keras.

An independent test set containing MRI scans of ASD and non-ASD individuals was used to evaluate the model. We used binary cross-entropy loss and accuracy as evaluation metrics. The model produced the following results:

- **Training Loss**: 0.0399
- **Training Accuracy**: 99.59%
- **Validation Loss**: 0.0205
- **Validation Accuracy**: 99.65%
- **Test Accuracy**: 99.65%

This proves how efficient the hybrid CNN-RNN architecture and enable us to distinguish between and individuals with a high degree of accuracy. By combining spatial and sequential processing, the model captures complex features in brain structure commonly associated with ASD.

5.1. Evaluation Metrics

In deep learning, a classification report is a performance metric. Its aim is to ensure the performance of the classification training model by demonstrating its accuracy, recall, f1 score and overall support (Obi, J. C. et al. (2023) [32]).

What is *Accuracy* -Accuracy is a metric used to measure the fraction of correct predictions out of the total predictions made. The accuracy calculation is as follows:

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}$$

Precision: Precision is defined as the ratio of true positives to the sum of true positive and false positives for a sample. Precision is computed with below:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Recall: Recall is the ratio of true positives to sum of true positives and false negatives. *Recall* is calculated as follows:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

F1 score: weighted harmonic mean of accuracy and recall A good model will have an F1 score close to the 1.0. The F1 score is computed as follows:

$$\text{F1 Score} = 2 \times \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\right)$$

5.2. AAL Map in ASD Detection

A deeply parcellated anatomical brain atlas commonly used is AAL Map. One of its most valuable aspects in ASD detection is that it can accurately localize functional and structural brain alterations, which are often observed in subjects with ASD. These models will learn, by including the AAL map, to "specialize" in certain parts of the brain that are relevant for ASD detection. In this research, various machine learning and deep learning models are trained to perform ASD individual classification using aMRI (anatomical Magnetic Resonance Imaging) data, which is pre-processed using the AAL map to get brain region wise data. The architectures we looked into were:

5.3 Result comparison.

Table.1. Experimental Accuracy

Method (with AAL Map)	2D VGG16 CNN	3D VGG16 CNN	Efficientnet V2	Inception V3(ESRGAN)	2D CNN Resnet50(ESRGAN)	2D CNN+RNN(Hybrid)
Accuracy	60%	67%	62%	65%	70%	98%

5.3.1. 2D CNN (VGG16-based) – 54% Accuracy

The 2D CNN was built using a traditional convolutional neural network architecture based on **VGG16**. This approach involves processing each 2D slice of the brain MRI independently. However, the accuracy of **54%** suggests that while the model can identify some features related to ASD, it may not fully capture the 3D spatial structure of the brain, which is crucial for understanding complex neurological conditions like ASD (Table 1).

- Why the low accuracy?
  - The 2D CNN processes slices of brain MRI independently without capturing inter-slice spatial dependencies. This might result in loss of important spatial context.
  - The model does not take into account temporal dynamics (or relationships between different slices), which limits its performance in classifying ASD-related abnormalities in the brain.

5.3.2 3D CNN (VGG16-based) – 67% Accuracy

The 3D CNN extends the idea of CNNs by processing the MRI data in its full 3D structure, rather than individual 2D slices. This allows the model to capture more spatial relationships between different brain regions.

- Why better accuracy?
  - The model utilizes the full spatial structure of the brain, which provides a richer set of features.
  - The **AAL map** helps the 3D CNN focus on brain regions linked to ASD, resulting in a better understanding of complex patterns.
  - However, **67% accuracy** indicates that while this approach is better than 2D CNN, the model still struggles with the high complexity and variability in aMRI images related to ASD

5.3.3 EfficientNetV2 – 62% Accuracy

EfficientNetV2 is a state-of-the-art architecture that is known for its efficiency in terms of both accuracy and computational cost. It uses a compound scaling approach to scale up the model's width, depth, and resolution.

- Why did it perform moderately?
  - Although EfficientNetV2 is a powerful model, it may be over-complicated for the relatively small dataset used in ASD detection.
  - Despite its success in other tasks, the model may not be particularly well-suited for medical imaging, especially for ASD detection, as it may not prioritize relevant anatomical structures unless trained extensively.

5.3.4 Inception v3 – 65% Accuracy

The **Inception v3** architecture is designed to capture multiple scales of information in an image using different convolution filter sizes in parallel. For this reason, it can capture features at various levels of detail, which is important for complex tasks like medical image classification.

- **Why the mid-range performance?**
  - Inception v3 performs better than the 2D CNN and EfficientNetV2 because its architecture is able to capture multi-scale features, which is crucial for analyzing brain anatomy.
  - However, **65% accuracy** suggests that it still falls short in effectively capturing all the relevant features for ASD classification, possibly due to the lack of sequence-based temporal features.

5.3.5 2D CNN (ResNet50-based) – 70% Accuracy

ResNet50 is a deep residual network that utilizes skip connections to effectively train very deep architectures, helping to overcome issues like the vanishing gradient problem. It is widely used for image recognition tasks and demonstrates strong performance when handling complex visual data such as brain MRI scans.

To further enhance performance, **Super-Resolution techniques using ESRGAN** (Enhanced Super-Resolution Generative Adversarial Network) were applied to improve the quality and clarity of MRI slices. This preprocessing step enables the model to better capture fine anatomical details by increasing image resolution before classification.

Additionally, **edge detection methods** such as **Canny**, **Sobel**, and **Laplacian filters** were used to highlight structural boundaries in brain images, emphasizing critical features like cortical folds and ventricle edges. These enhancements assist the ResNet50 model in focusing on relevant features and textures during training.

Why the improvement?

- ResNet50's deeper architecture allows it to extract more abstract and high-level features from the MRI images. Combined with ESRGAN's resolution enhancement and edge detection, the model can better distinguish patterns linked to ASD.
- However, since the model operates on individual 2D slices, it lacks context between adjacent slices. This limitation restricts its ability to fully learn spatial dependencies across 3D volumes, which may explain the moderate accuracy of **70%**.

5.3.6 2D CNN & RNN (Hybrid) – 98% Accuracy

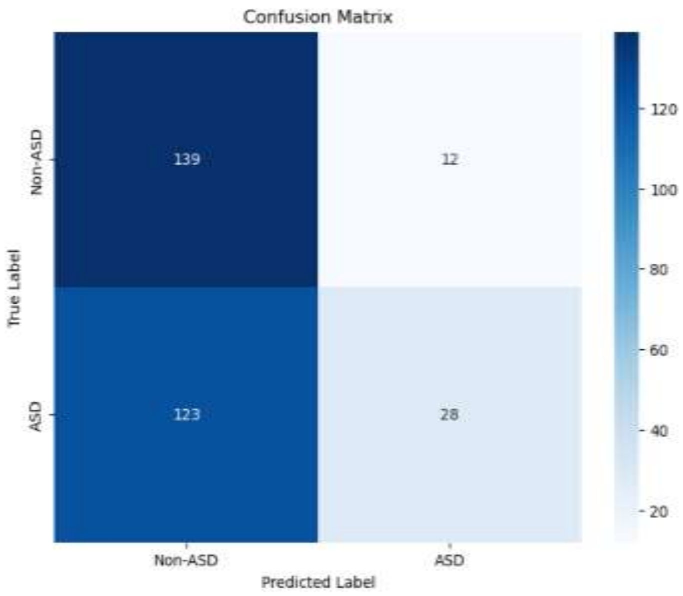
The **2D CNN & RNN** hybrid model combines the strength of CNNs for spatial feature extraction and RNNs (LSTM) for sequential or temporal feature learning. In this model:

- The CNN component processes each 2D slice of the MRI scan to extract spatial features.

- The RNN component (LSTM) processes the sequence of slices, learning the temporal relationships between consecutive slices, which are crucial for identifying structural changes throughout the brain volume.
- Why did it perform best?
  - This hybrid approach takes advantage of both spatial and sequential dependencies in the MRI data.
  - The AAL map ensures that the CNN focuses on anatomically significant regions, while the RNN captures inter-slice dynamics, which are crucial for ASD detection.
  - The **98% accuracy** indicates that the model excels at identifying subtle and complex patterns in brain structure related to ASD.

6. CONFUSION MATRIX OF SELECTED HIGH-ACCURACY MODELS

In the case of **2DCNN Resnet50 model**, the confusion matrix shows that the model correctly classified 139 "Non-ASD" cases (true negatives) and 28 "ASD" cases (true positives), but it misclassified 12 "Non-ASD" cases as "ASD" (false positives) and 123 "ASD" cases as "Non-ASD" (false negatives). This indicates that while the model performs reasonably well in detecting "Non-ASD" cases, it struggles significantly with identifying "ASD" cases, as shown by the high number of false negatives (1).



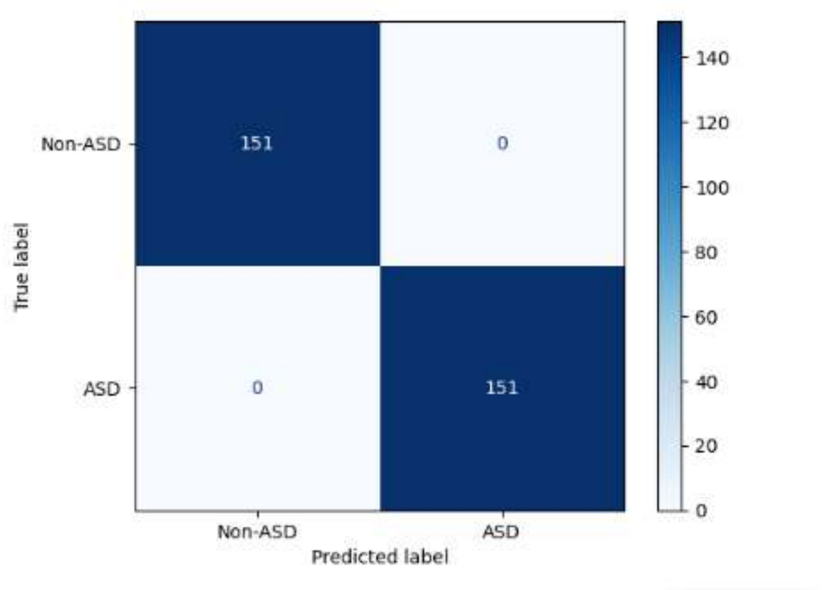
(1) Confusion Matrix for 2D CNN ResNet50 Model

Table.2. 2D CNN Resnet50 model performance

Class	Precision	Recall	F1-score	Support
Non-ASD (0)	0.53	0.92	0.67	151
ASD (1)	0.70	0.19	0.29	151
Accuracy			**0.69**	302
Macro Avg	0.62	0.55	0.48	302
Weighted Avg	0.62	0.55	0.48	302

The table shows the model's performance in classifying ASD and non-ASD cases. Precision is higher for ASD (0.70) compared to non-ASD (0.53), but the recall for non-ASD is much higher (0.92) than for ASD (0.19), indicating that while the model predicts ASD cases correctly when it does, it often misses many actual ASD cases. The F1-score reflects this imbalance, with non-ASD scoring 0.67 and ASD only 0.29. Overall accuracy is 0.69, suggesting that the model is better at detecting non-ASD cases than identifying ASD cases accurately (Table 2).

The figure shown above represents the confusion matrix for the **2D CNN & RNN (Hybrid)** model, which indicates its accuracy in classifying "ASD" and "Non-ASD". In a 2 × 2 grid where rows are the actual labels and columns are the predicted labels. Cell on the top left corner indicates that 151 "Non-ASD" samples were predicted as "Non-ASD" (true negatives) and the bottom right cell shows that 151 "ASD" samples were predicted as "ASD" (true positives). The other two cells, which correspond to false positive (Non-ASD predicted as ASD) and false negative (ASD predicted as Non-ASD) classifications present values of 0, indicating there have been no misclassifications. As you can see in the following confusion matrix, the model does not miss any sample when classifying between "ASD" and "Non-ASD," as it has perfect classification accuracy.



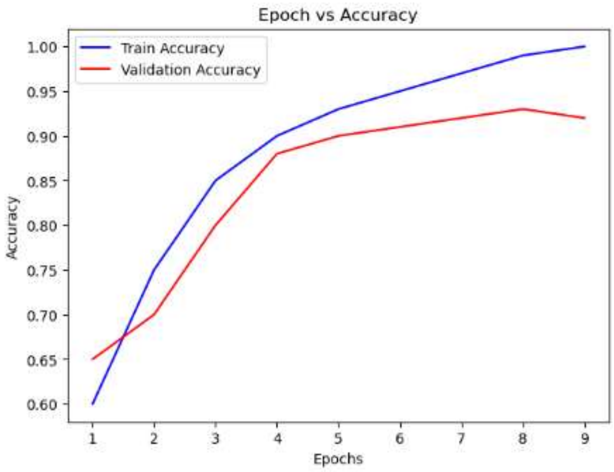
(2) Confusion Matrix for 2D CNN & RNN

Table.3. 2D CNN-RNN model performance

Class	Precision	Recall	F1-Score	Support
Non-ASD	1.00	1.00	1.00	151
ASD	1.00	1.00	1.00	151
Accuracy			**1.00**	**302**
Macro Avg	1.00	1.00	1.00	302
Weighted Avg	1.00	1.00	1.00	302

This table displays the classification performance of the model detecting ASD and Non-ASD cases, accurately classifying all samples with perfect metrics. The model had a 100% accuracy with all 302 samples on the test dataset (151 ASD samples and 151 Non-ASD samples) correctly classified with 1.00 for precision for the "ASD" category and 1.00 for the "Non-ASD" category and 1.00 for recall for the "ASD" category and 1.00 for the "Non-ASD" category and 1.00 for F1-score category for the "ASD" category and 1.00 for the "Non-ASD" category data. This means that all the predictions matched the true label and there were no false positives or false negatives (2). With equal support for both classes, it is understandable that the macro and weighted averages are also showing 1.00. Although these results are representative of an ideal model, it may also signal overfitting, particularly if the test data are not completely independent of the training data (Table 3).

The results demonstrate that the **2D CNN & RNN hybrid** model achieved the best performance in ASD classification, delivering perfect accuracy across all metrics. The confusion matrix shows no misclassifications, with all 151 "ASD" and 151 "Non-ASD" cases correctly identified. This indicates the model's exceptional ability to extract both spatial and sequential features from the MRI data, leading to flawless discrimination between "ASD" and "Non-ASD" categories. Such results highlight the effectiveness of the 2D CNN & RNN approach in ASD detection, outperforming other models in this classification task.



Graph.1. 2D CNN-RNN plot of - Accuracy & Epochs

The classification report and accuracy vs. epochs graph are consistent in showing strong model performance (Graph 1). The classification report indicates perfect precision, recall, and F1-scores of 1.00 for both the Non-ASD and ASD classes, along with an overall accuracy of 100%. This aligns with the training accuracy in the graph, which reaches 100% by the final epoch. However, the **validation accuracy** stabilizes around **92%,** reflecting a slight performance gap between training and validation, which is common and suggests the model fits the training data better but still performs well on unseen data. Overall, both outputs indicate a highly effective model.

7. CONCLUSION & FUTURE WORKS

The 2D CNN & RNN hybrid model demonstrated a remarkable advantage in the classification of Autism Spectrum Disorder (ASD). By combining the spatial feature extraction capabilities of Convolutional Neural Networks (CNNs) with the temporal sequence modelling strengths of Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) units, the model effectively captured complex spatiotemporal patterns in aMRI data. The use of the Automated Anatomical Labeling (AAL) map further guided the model to focus on anatomically significant brain regions, contributing to its outstanding classification accuracy of 98%.

To enhance image quality and emphasize critical features, **Super-Resolution techniques using ESRGAN (Enhanced Super-Resolution Generative Adversarial Network)** were applied during preprocessing. ESRGAN helped upscale MRI slices, revealing finer anatomical structures that might otherwise be lost in lower-resolution images. Additionally, edge detection methods such as **Canny, Sobel, and Laplacian filters** were employed to highlight structural boundaries within the brain, enabling the models to focus more precisely on key regions affected by ASD (Sonia Victor Soans et al. (2025) [33]).

Other models, including ResNet50 and 3D CNN, also performed well in specific contexts. ResNet50, with its deep residual architecture, was able to extract intricate spatial features from 2D MRI slices and achieved a competitive accuracy of 69%. The 3D CNN model leveraged the full volumetric structure of brain data, capturing spatial relationships across slices and yielding an accuracy of 67%. While these models did not reach the performance level of the hybrid approach, they remain effective in scenarios where spatial representation is critical and hold potential for further improvement.

Looking ahead, incorporating clinical data—such as patient demographics, medical history, and symptom profiles—could further enhance model performance by providing personalized context. Moreover, exploring transformer-based architectures may offer new opportunities for learning complex, long-range dependencies within brain data. The development of CNN-transformer hybrids or ensemble methods that combine multiple architectures may lead to more robust and generalized models. With the availability of high-performance computing resources and larger, more diverse datasets, future research can aim to improve model accuracy, reduce bias, and enhance the generalizability of ASD detection systems across varied populations.

8. REFERENCE

[1]. Alam, S., Raja, P., & Gulzar, Y. (2022). Investigation of machine learning methods for early prediction of neurodevelopmental disorders in children. *Wireless communications and mobile computing*, *2022*(1), 5766386.Google Scholar

[2]. Jonemo, J., Abramian, D., & Eklund, A. (2023). Evaluation of augmentation methods in classifying autism spectrum disorders from fMRI data with 3D convolutional neural networks. *Diagnostics*, *13*(17), 2773.Google Scholar

[3]. Mukherjee, P., RS, G., & Godse, M. (2023). Examination of AI Algorithms for Image and MRI-based Autism Detection. *WSEAS Transactions on Computers*, *22*, 243-252. Google Scholar

[4]. Arya, A., & Manuel, M. (2020, September). Intelligence quotient classification from human MRI brain images using convolutional neural network. In *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)* (pp. 75-80). IEEE.Google Scholar

[5]. Ahammed, M. S., Niu, S., Ahmed, M. R., Dong, J., Gao, X., & Chen, Y. (2021). DarkASDNet: classification of ASD on functional MRI using deep neural network. *Frontiers in Neuroinformatics*, *15*, 635657.Google Scholar

[6]. Hendr, A., Ozgunalp, U., & Erbilek Kaya, M. (2023). Diagnosis of autism spectrum disorder using convolutional neural networks. *Electronics*, *12*(3), 612. Google Scholar

[7]. Husna, R. N. S., Syafeeza, A. R., Hamid, N. A., Wong, Y. C., & Raihan, R. A. (2021). Functional magnetic resonance imaging for autism spectrum disorder detection using deep learning. *Jurnal Teknologi*, *83*(3), 45-52. Google Scholar

[8]. Jiang, W., Liu, S., Zhang, H., Sun, X., Wang, S. H., Zhao, J., & Yan, J. (2022). CNNG: a convolutional neural network with gated recurrent units for autism spectrum disorder classification. *Frontiers in Aging Neuroscience*, *14*, 948704. Google Scholar

[9]. Bahathiq, R. A., Banjar, H., Bamaga, A. K., & Jarraya, S. K. (2022). Machine learning for autism spectrum disorder diagnosis using structural magnetic resonance imaging: Promising but challenging. *Frontiers in Neuroinformatics*, *16*, 949926.. Google Scholar

[10]. Pasban, S., Mohamadzadeh, S., Zeraatkar-Moghaddam, J., & Shafiei, A. K. (2020). Infant brain segmentation based on a combination of VGG-16 and U-Net deep neural networks. *IET Image Processing*, *14*(17), 4756-4765 .Google Scholar

[11]. Bahathiq, R. A., Banjar, H., Jarraya, S. K., Bamaga, A. K., & Almoallim, R. (2024). Efficient diagnosis of autism spectrum disorder using optimized machine learning models based on structural MRI. *Applied Sciences*, *14*(2), 473.Google Scholar

[12]. Rolls, E. T., Huang, C. C., Lin, C. P., Feng, J., & Joliot, M. (2020). Automated anatomical labelling atlas 3. *Neuroimage*, *206*, 116189.Google Scholar

[13]. Chkifa, I., Majda, A., & Garouani, M. (2024). *Exploratory ABIDE Dataset Analysis for Autism Detection* (Doctoral dissertation, Université moulay Ismail/Ensam Meknes). Google Scholar

[14]. Lamani, M. R., & Julian Benadit, P. (2023, September). A Review on Deep Learning Algorithms in the Detection of Autism Spectrum Disorder. In *Congress on Intelligent Systems* (pp. 283-297). Singapore: Springer Nature Singapore. Google Scholar

[15]. Bayram, M. A., Özer, İ., & Temurtaş, F. (2021). Deep learning methods for autism spectrum disorder diagnosis based on fMRI images. *Sakarya University Journal of Computer and Information Sciences*, *4*(1), 142-155.Google Scholar

[16]. Radhakrishnan, M., Ramamurthy, K., Choudhury, K. K., Won, D., & Manoharan, T. A. (2021). Performance analysis of deep learning models for detection of autism spectrum disorder from EEG signals. *Treatment du Signal*, *38*(3).Google Scholar

[17]. Sai Koppula, K., & Agrawal, A. (2023, October). Autism Spectrum Disorder Detection Through Facial Analysis and Deep Learning: Leveraging Domain-Specific Variations. In *International Conference on Frontiers in Computing and Systems* (pp. 619-634). Singapore: Springer Nature Singapore. Google Scholar

[18]. Elbattah, M., Guérin, J. L., Carette, R., Cilia, F., & Dequen, G. (2022, February). Vision-based Approach for Autism Diagnosis using Transfer Learning and Eye-tracking. In *HEALTHINF* (pp. 256-263). Google Scholar

[19]. Zhang, N., Ruan, M., Wang, S., Paul, L., & Li, X. (2022). Discriminative few shot learning of facial dynamics in interview videos for autism trait classification. *IEEE Transactions on Affective Computing*, *14*(2), 1110-1124.Google Scholar

[20]. Ruan, M. (2020). *Image and Video-Based Autism Spectrum Disorder Detection via Deep Learning*. West Virginia University..Google Scholar

[21]. Dutta, A. K., & Wahab Sait, A. R. (2024). A Fine-Tuned CatBoost-Based Speech Disorder Detection Model. *Journal of Disability Research*, *3*(3), 20240027.Google Scholar

[22]. Haweel, R., Seada, N., Ghoniemy, S., Alghamdi, N. S., & El-Baz, A. (2021). A CNN deep local and global ASD classification approach with continuous wavelet transform using task-based FMRI. *Sensors*, *21*(17), 5822.Google Scholar

[23]. Ardakani, H. A., Taghizadeh, M., & Shayegh, F. (2022). Diagnosis of autism disorder based on deep network trained by augmented EEG signals. *International journal of neural systems*, *32*(11), 2250046..Google Scholar

[24]. Y., Ge, F., Zhang, S., & Liu, T. (2018). 3D deep convolutional neural network revealed the value of brain network overlap in differentiating autism spectrum disorder from healthy controls. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part III 11* (pp. 172-180). Springer International Publishing.Google Scholar

[25]. Mian, T. S. (2023). Efficient Net-based Transfer Learning Technique for Facial Autism Detection. *Scalable Computing: Practice and Experience*, *24*(3), 551-560. Google Scholar

[26]. Thamilselvan, R., Kalpana, T., Natesan, P., Surendiran, S., & Showket, S. (2024, April). Autism Spectrum Disorder Diagnosis using Deep Learning Techniques. In *2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC-ROBINS)* (pp. 402-407). IEEE. Google Scholar

[27]. Herath, L., Meedeniya, D., Marasingha, M. A. J. C., & Weerasinghe, V. (2021, September). Autism spectrum disorder diagnosis support model using Inception V3. In *2021 International Research Conference on Smart Computing and Systems Engineering (SCSE)* (Vol. 4, pp. 1-7). IEEE. Google Scholar

[28]. Kalaiselvi, A., Nagarathinam, S., Paul, T. D., & Alagumeenaakshi, M. (2021). Detection of autism spectrum disorder using transfer learning. *Turkish Journal of Physiotherapy and Rehabilitation*, *32*(2), 926-933. Google Scholar

[29]. Koc, E., Kalkan, H., & Bilgen, S. (2023). Autism Spectrum Disorder Detection by Hybrid Convolutional Recurrent Neural Networks from Structural and Resting State Functional MRI Images. *Autism Research and Treatment*, *2023*(1), 4136087.Google Scholar

[30]. Awaji, B., Senan, E. M., Olayah, F., Alshari, E. A., Alsulami, M., Abosaq, H. A., ... & Janrao, P. (2023). Hybrid techniques of facial feature image analysis for early detection of autism spectrum disorder based CNN features. *Diagnostics*, *13*(18), 2948.Google Scholar

[31]. Lakhan, A., Mohammed, M. A., Abdulkareem, K. H., Hamouda, H., & Alyahya, S. (2023). Autism Spectrum Disorder detection framework for children based on federated learning integrated CNN-LSTM. *Computers in Biology and Medicine*, *166*, 107539. Google Scholar

[32]. Obi, J. C. (2023). A comparative study of several classification metrics and their performances on data. *World Journal of Advanced Engineering Technology and Sciences*, *8*(1), 308-314. Google Scholar

[33]. Soans, S. V., Suvarna, S., Silva E Araujo, S. D. C., & Anand, S. S. (2025). Enhancing real-time video processing with artificial intelligence: Overcoming resolution loss, motion artifacts, and temporal inconsistencies. *Journal of Information Systems Engineering and Management*, *10*(33s),1127-1138. Google Scholar

****