Received: 11 August 2023 | Revised: 8 December 2023 | Accepted: 10 January 2024

IET Computer Vision

DOI: 10.1049/cvi2.12271

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

ASDNet: A robust involution-based architecture for diagnosis of autism spectrum disorder utilising eye-tracking technology

Nasirul Mumenin[1] | Mohammad Abu Yousuf[2] | Md Asif Nashiry[3] | A. K. M. Azad[4] | Salem A. Alyami[5] | Pietro Lio'[6] | Mohammad Ali Moni[7,8]

[1]Department of Information and Communication Technology, Bangladesh University of Professionals, Dhaka, Bangladesh

[2]Institute of Information Technology, Jahangirnagar University, Savar, Bangladesh

[3]Department of Data Analytics, Northern Alberta Institute of Technology, Edmonton, Alberta, Canada

[4]Department of Mathematics and Statistics, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia

[5]Department of Mathematics and Statistics, College of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia

[6]Department of Computer Science and Technology, The University of Cambridge, Cambridgeshire, UK

[7]Centre for AI & Digital Health Technology, Charles Sturt University, AI & Cyber Future Institute, Orange, New South Wales, Australia

[8]Rural Health Research Institute, Charles Sturt University, Orange, NSW, Australia

Correspondence
Mohammad Abu Yousuf, Salem A. Alyami and Mohammad Ali Moni.
Email: yousuf@juniv.edu, saalyami@imamu.edu.sa and mmoni@csu.edu.au

Funding information
Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU), Grant/Award Number: IMSIU-RP23004

Abstract

Autism Spectrum Disorder (ASD) is a chronic condition characterised by impairments in social interaction and communication. Early detection of ASD is desired, and there exists a demand for the development of diagnostic aids to facilitate this. A lightweight Involutional Neural Network (INN) architecture has been developed to diagnose ASD. The model follows a simpler architectural design and has less number of parameters than the state-of-the-art (SOTA) image classification models, requiring lower computational resources. The proposed model is trained to detect ASD from eye-tracking scanpath (SP), heatmap (HM), and fixation map (FM) images. Monte Carlo Dropout has been applied to the model to perform an uncertainty analysis and ensure the effectiveness of the output provided by the proposed INN model. The model has been trained and evaluated using two publicly accessible datasets. From the experiment, it is seen that the model has achieved 98.12% accuracy, 96.83% accuracy, and 97.61% accuracy on SP, FM, and HM, respectively, which outperforms the current SOTA image classification models and other existing works conducted on this topic.

KEYWORDS
autism spectrum disorder, convolutional neural network, eye tracking, involutional neural network, Monte Carlo Dropout, uncertainty analysis

1 | INTRODUCTION

Autism spectrum disorder (ASD), commonly referred to as autism, is a widespread, highly heritable, diverse neurodevelopmental condition with cognitive underpinnings and often co-occuring with other disorders [1]. ASD is a diversified impairment, and indicating this diversity, the term autism has been adopted in a variety of forms to designate both a broader manifestation and a specific diagnosis since its classification as a subset of pervasive developmental disorders. Autism is characterised by difficulties with interpersonal interaction and engagement, sensorial abnormalities, repetitive actions, and varying degrees of intellectual limitation. Figure 1 depicts the symptoms of ASD [2]. ASD is one of the most common forms

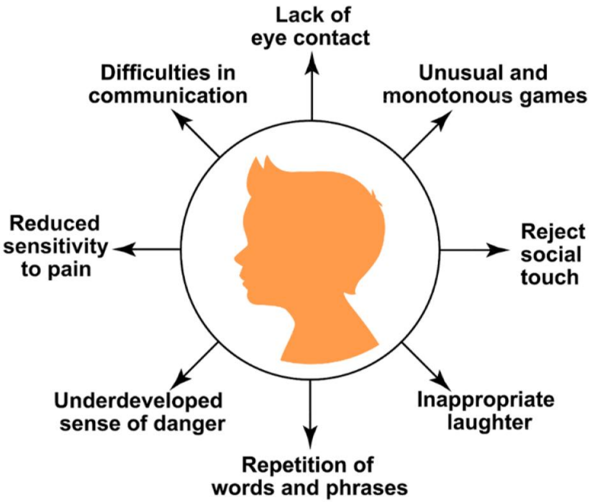
Lack of eye contact
Difficulties in communication
Unusual and monotonous games
Reduced sensitivity to pain
Reject social touch
Underdeveloped sense of danger
Inappropriate laughter
Repetition of words and phrases

FIGURE 1 Symptoms shown by a person with autism spectrum disorder.

of impairment worldwide. The current estimation shows that one in a hundred children are affected by it globally. ASD has a frequency of about under 1% globally, although estimates are higher in high-income nations [3]. Boys are over four times more likely to have ASD than girls. During the 2009–2017 research period, about 1 in 6 (17%) children who are aged 3–17 years were diagnosed with a developmental impairment, as per the report by their parents. Among these conditions were ASD, attention-deficit disorder, cerebral palsy, etc. [4]. This estimation is typical, although reported occurrence varies widely between research studies. However, several rigorous investigations have revealed noticeably higher numbers. Unknown is the frequency of ASD in several low and middle-income nations.

Currently, there is no treatment for ASD that will result in a full recovery [5]. Nevertheless, early diagnosis and treatment can alleviate the situation. Therefore, autistic people require adequate support and consideration [6]. Due to the complexities, diversity, and uniformity of ASD, medical guidance documents recommend multidisciplinary teams for ASD diagnosis [7, 8]. The most sensitive and specific diagnostic tests are the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R) [9]. Despite the fact that these tools are effective at identifying ASD-related behaviours, their 'diagnostic discrimination and required resources' have been criticised by numerous users. Specifically, the use of ADOS and ADI-R costs a significant amount of time and can contribute to the premature diagnosis of ASD [10]. Accessible, efficient, and more effective detection methods, particularly for children, are an urgent requirement to resolve these issues. Several biomarkers have gained recognition for their ability to identify individuals possessing an elevated risk of being affected by ASD. Prior to the onset of behavioural symptoms, these biomarkers evaluate genetical differences, early brain structural and functional connectivity, visual orientation etc. Using some tools that include functional magnetic resonance imaging [11, 12], electroencephalography (EEG) [13, 14], metabolic disorder testing [15, 16], and facial expression analysis [17, 18], it is possible to diagnose ASD in early ages. Eye-tracking has proven its capability as a new tool in the detection of ASD in recent times [19–21]. Eye-tracking (ET) is a method that can be utilised to capture, track, and analyse the eye movements or absolute point of gaze, which denotes the location in the visual scene where the eye gaze is directing attention [22, 23]. The ET technology seeks to utilise the perceptual aspects of ASD as well as recognise the atypical visual attention that leads to ASD's clinical manifestations [24]. ET has proven to be a highly effective diagnostic tool due to the cause that aberrant fixation movements are one of the defining characteristics of ASD. ET can be translated into scanpath or saliency maps for the enhancement of visualisation. Therefore, ET can be used to identify ASD characteristics and enhance diagnostic precision [25, 26]. Due to the complexity of the undertaking, artificial intelligence, specifically Deep Learning (DL), has been incorporated to enhance diagnostic accuracy.

Given the importance of the issue, a substantial amount of research has been conducted over the years on the diagnosis of ASD. Some social and behavioural symptoms, such as lack of attention, eye contact, and social interaction, may be more challenging to recognise in the early stage. Therefore, conventional methods are insufficient for early diagnosis, which hinders recovery. Due to these differing levels of symptom severity, the detection is quite complex and requires further investigation. Machine Learning (ML) and DL incorporated with ET can play a vital role in bringing a promising solution in this case.

Much research has been conducted to find the association of ET with the diagnosis of ASD. Bataineh et al. [27] provided an eye-tracking analysis research to assist and develop an understanding of the visual character and movement of children with ASD and Typically Developed (TD) while seeing a socially rich stimulus consisting of social interactions. The researchers discovered a substantial difference in the viewing patterns and behaviours of the two groups when presented with a scenario including human and social interaction components. The research also demonstrates that a considerable proportion of autistic individuals had little interest in and time spent focusing on the face region, as seen by their extensive fixation on non-facial areas, which correlates with a lack of interest in socially important information. Solovyova et al. [28] examined novel ways for online autism diagnoses and created an algorithm that can predict the likelihood of ASD based on the child's gaze activity. Experimental findings supported the notion that ET is effective for the early identification of eye-movement characteristics that may serve as ASD indicators. Eraslan et al. [29] examined if it was feasible to predict autism based on eye-movement sequences with sufficient accuracy to be independent of individual online sites. The authors discovered that sequential data analysis yields more consistent findings than non-sequential data analysis, which might aid in overcoming disadvantages in stimulus selection. Mazumdar et al. [30] researched the early detection of ASD in children. Their strategy is based on analysing children's visual behaviour when investigating pictures. Based on picture content, fixations, and centre bias, scanpaths were analysed to derive features capable of detecting atypical viewing behaviour. Cilia et al. [31] explored ET as an integrated element of ASD screening on the basis of typical components of eye gazing. This work contributes to the growing body of research on ET technologies to facilitate ASD screening. To categorise individuals with ASD, the suggested technique used ET with visualisation and CNN. Almourad et al. [32] presented an ET analysis investigation in order to comprehend the visual behaviour of children with TD and ASD when watching human face stimuli. In comparison to children without ASD, a considerable proportion of ASD participants exhibited little interest in and fixation on the face region, as shown by significant time spent fixating on non-facial locations.

Several research studies have been conducted on the detection of ASD using ML algorithms. Using a number of ML approaches, Akter et al. [33] analysed eye-gazing photos to diagnose autism. As the scan path pictures obtained by the ASD and TD groups were almost identical, the k-means clustering approach was employed to construct four clusters. Using Artificial Neural Network (ANN), they achieved 87% accuracy. Shihab et al. [34] concentrated on evaluating the dataset of people of various ages with ASD using the PCA approach. Their primary objective was to adopt PCA to minimise the dimension of the data and maintain just the characteristics that give discriminating patterns. In the case of adults, their classification studies yielded a sensitivity of 78.6% and a specificity of 82.47%, but for children, the sensitivity was 87.5%, and the specificity was 95.7%. Carette et al. [35] provided a method for assisting with ASD diagnosis, with a special emphasis on young infants. The primary concept was to convert ET scanpaths into a visual presentation; hence, detection may be handled as an image classification issue. Experiments included many ML approaches and an ANN. Using basic ANN, they attained AUC >90% [36]. According to them, providing a proactive concept of further testing may be useful to explore a kid exhibiting symptoms of this illness. Akter et al. [37] proposed an ML-based approach for early-stage detection. Gathering the early detection data of children, adults, toddlers and adolescents, they transformed and applied several ML techniques for classification. Uddin et al. [38] developed an ML-based framework considering the ASD screening dataset for toddlers for the detection of ASD. They applied the SMOTE method to balance the dataset and applied several ML techniques among which AdaBoost got the highest accuracy. A method has been proposed for the identification of subgroups of ASD by Lin et al [39]. They used ML to analyse the microarray data to identify groups having similar gene expression profiles. They achieved the highest accuracy using RF and SVM. Kanhirakadavath and Chandran [40] evaluated the use of ET data in the early diagnosis of autism in children using ML techniques. The authors investigated the efficacy of several ML approaches to determine the best model for the detection of ASD using scanpath pictures from ET. The suggested DNN model outperforms conventional ML techniques with a 97% AUC.

Many researchers have utilised DL and transfer learning for the detection of ASD. Based on 700 photos and related eye movement patterns of ASD and TD, Xie et al. [41] constructed a unique two-stream DL network for this recognition. Their proposed model achieved 0.95 accuracy. They defined contributions to categorisation at the level of a single picture and non-linear integration of information at this level during classification. Using ET data, Elbattah et al. [42] suggested a sequence learning method for identifying autism. Their primary concept was to portray ET data as textual strings that present the fixations and saccade patterns. The data was then categorised using CNN and LSTM to identify ASD and TD. Mahalakshmi and Praveena [43] suggested a technique for diagnosing ASD and TD using CNN for the fixation maps of the respective observer's gaze at a given picture. Their suggested CNN model obtains a validation accuracy of 75.23%. Wei et al. [44] suggested an image-level approach to determine if a scanpath belongs to a kid with ASD or a youngster with typical development. The seen picture is first encoded using a visual feature encoder. For classification, an LSTM-based model was combined with embedding and dynamic filters. The data is then utilised to build an ML classifier with an accuracy of 75%. Duan et al. [45] created a database, which comprises 500 pictures and related ET data. They examined the performance of five NN-based saliency estimation techniques with the original and the fine-tuned models that were created by them. Fawaz Waseelallah Alsaade and Mohammed Saeed Alzahrani [46] suggested a technique for identifying autism based on facial characteristics using a simple online application utilising a DL algorithm. CNN used transfer learning and the Flask framework. Pretrained models Xception, VGG19, and NASNETMobile were utilised. The dataset utilised to evaluate these models consists of 2940 face photos taken from the Kaggle platform. Their Xception model obtained 91% accuracy, followed by VGG19 (80%) and NASNETMobile (78%). The review presents that there are still various scopes of work and scopes for improvement. We have found no such work where various types of ET images, such as scanpath, saliency map, and fixation map, have been combined to diagnose ASD. In this work, we have addressed the limitations found in the literature.

The main factor on which this research is conducted is that people with ASD have an uncommon focus pattern, which can be identified early by analysing images created using ET. Several types of research have proven that ET, visual attention, and saliency can be used for the diagnosis of autism and can be an important biomarker for early detection [47–52]. In the existing studies, the researchers have worked with only one type of ET data to diagnose autism, mostly SP or FM or HM images. This creates a limitation for the model to have the capability of generalisation. The models used in previous works are mostly based on CNN techniques, and they have a massive number of parameters that take a huge number of images to train properly and require much computational power. These existing approaches require sophisticated hardware to be executed because of the numerous parameters they demand. In this study, a lightweight deep Involutional Neural Network (INN) model has been developed to classify people with ASD from TD or healthy control (HC). The proposed model consists of a small number of parameters comparatively. The model has been tested with three different types of ET images to diagnose ASD from any ET images. The diagnosis procedure can be efficiently automated using this model. So, the proposed INN model can be utilised to create an automatic and effective detection system for ASD. The main contribution of our research is that we have built a lightweight INN-based DL model for the diagnosis of ASD. The model consists of relatively a small number of layers and parameters; thus, it can be called a lightweight model. Three different types of ET images have been used to train and evaluate the model. And finally, we have performed an uncertainty analysis to ensure the certainty of the output and increase the model's performance. Model validation has been done by comparing with other state-of-the-art image classification models and other existing literature.

2 | PROPOSED APPROACH

The proposed approach consists of some key steps. Figure 2 depicts the key steps of the proposed methodology. Firstly, we have collected a dataset for our work. After that, the data has been preprocessed. Then, the INN model has been created, which receives the preprocessed input images. The model has been trained using the training dataset and then tested using the test dataset. The performance of the INN model is then compared with existing works to validate its performance. Lastly, uncertainty analysis is performed to prove the effectiveness and increase the model performance.

2.1 | Dataset

We have utilised two publicly available datasets for this work (ASD Dataset 1 [53] and ASD Dataset 2 [54]). These two datasets have been addressed as Dataset-1 and Dataset-2, respectively, throughout this paper.

2.1.1 | Dataset-1

This dataset has been created by Carette et al. [53]. One of their key goals was to identify autism early; therefore, they restricted the participants' ages to those between 3 and 13 years old. This dataset consists of 59 samples which include 30 ASD-infected children and 29 controls. Among 59 children, 38 were boys, and 21 were girls. The SMI RED mobile [55] served as the main equipment for ET investigations. A participant's eye movement across the screen was represented by coordinates in the raw ET data. The coordinates were then used to produce the visualisation. The entire process of data transformation was implemented in Python. The dataset contains ET Scanpath (SP) images of ASD and HC persons. Figure 3 shows some sample SP images from this dataset.

2.1.2 | Dataset-2

This dataset has been created for the challenge 'Saliency4ASD' [56] and made publicly available in Ref. [54]. It comprises 300 natural scene photos and ET data gathered from 14 ASD patients and 14 HCs. The natural scene photos were used as

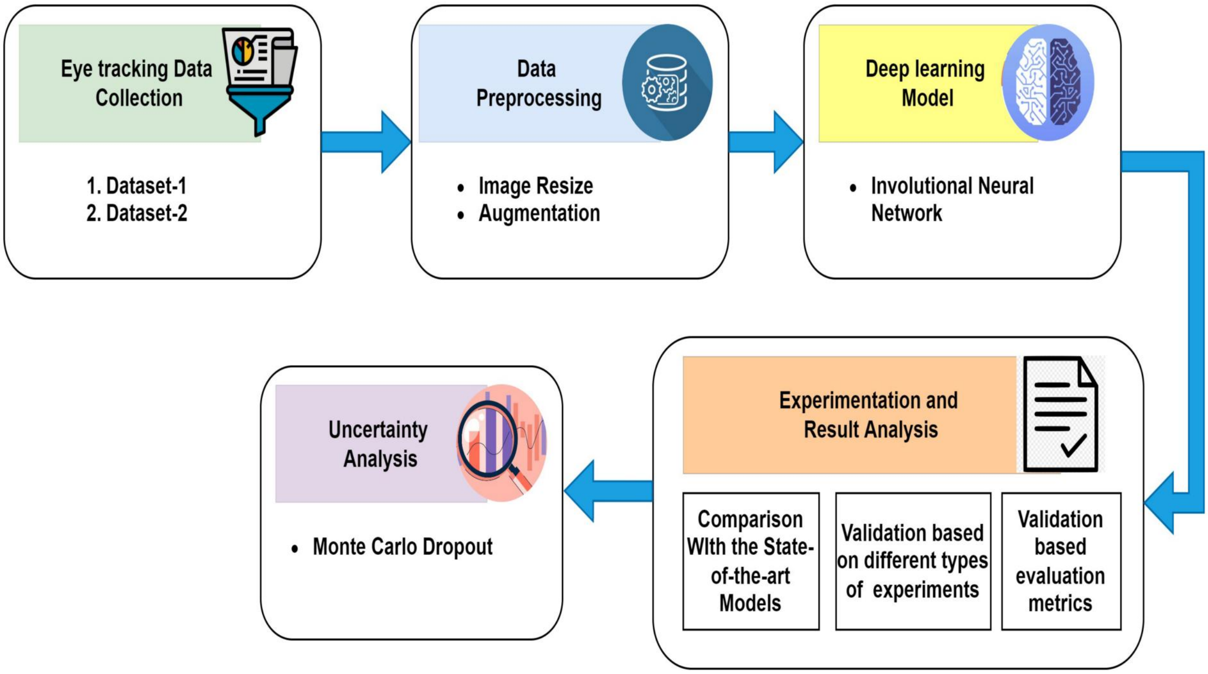
Eye tracking Data Collection
1. Dataset-1
2. Dataset-2
Data Preprocessing
• Image Resize
• Augmentation
Deep learning Model
• Involutional Neural Network
Experimentation and Result Analysis
Comparison WIth the State-of-the-art Models
Validation based on different types of experiments
Validation based evaluation metrics
Uncertainty Analysis
• Monte Carlo Dropout

FIGURE 2 Workflow of the proposed method. Datasets have been gathered initially. Then, the data has been scaled, and augmentation has been applied. Next, the involutional neural network model has been constructed and assessed through various experiments. Lastly, uncertainty analysis has been employed to measure output certainty.

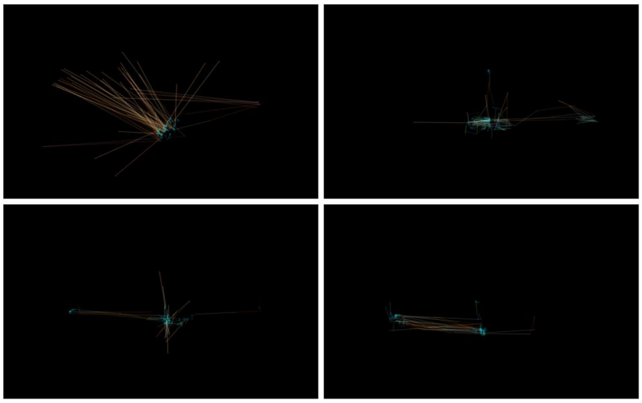

FIGURE 3 Sample of eye tracking SP images from Dataset-1. The SP presents the eye movement of the subject over a time period.

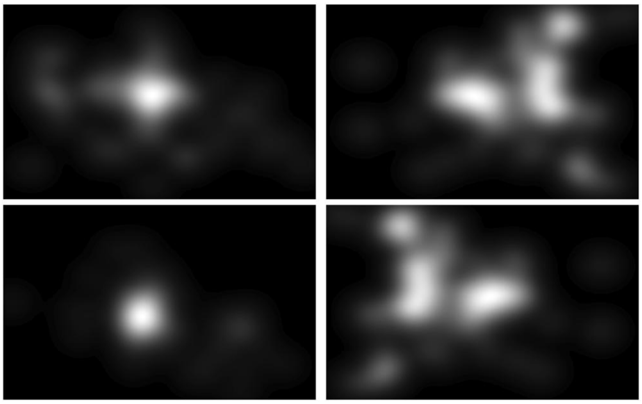

FIGURE 4 Sample of FM images from Dataset-2. The FM presents the saliency or focus of the subject on the screen over a time period through B&W colour. The brightness denotes the amount of time a subject spends on a particular point.

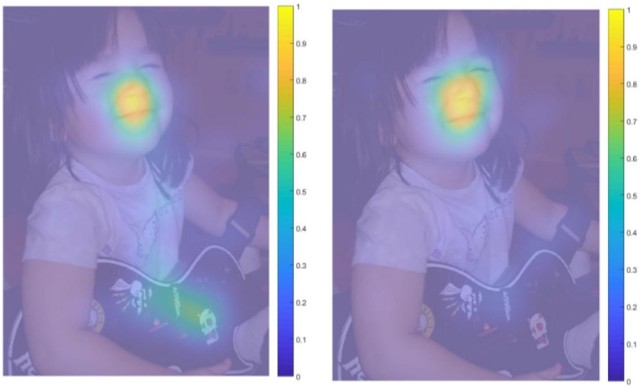

FIGURE 5 Sample of HM images from Dataset-2. The HM presents the saliency or focus of the subject on the screen over a time period through RGB colour. The time a subject spends on a particular point is denoted through colour variation. The colour variation is the amount of time spent on a particular point. Blue denotes a small amount of time, green denotes a medium amount, and yellow denotes a long amount of time.

stimuli. The dataset contains ET data and 300 photos that were obtained from it. The dataset includes pictures of stimuli, Fixation Maps (FM), and Heatmaps (HM). Using a Tobii T120 Eye Tracker, they exhibited the images and recorded the participants' eye movements. The device's sampling rate was set to 120 Hz, and its practical effective range is between 50 and 80 cm. The participants were 8 years old on average. Figure 4 shows some sample FM images, while Figure 5 shows some sample HM images from this dataset.

2.2 | Preprocessing

Following data collection, images have been separated into training, validation, and test dataset while maintaining the proportion of classes for people with ASD against those without ASD in random order. The proportions for training, validation and testing datasets are 80%, 10%, and 10%, respectively. The images in both datasets have been preprocessed by boosting the brightness of the images and scaling to (32 × 32). The brightness of the images has been increased by 5%–10%. Experiments have shown that the increase in brightness allows the model to perform better.

We have applied augmentation to produce pictures with different viewing modifications. Following the process of augmentation, 2933 additional photos in Dataset-1 and 2541 additional images in Dataset-2 were added to the collections. The Keras package [57] offers an API that is easy to use for data augmentation that considerably simplifies the augmentation method. We used three distinct augmentation approaches in our process: rotation, flipping, and scaling. The Imagedatagenerator technique flips and rotates pictures at random. The value of scaling has been set to 5%–15%.

2.3 | Proposed involutional neural network

CNN is a special type of DL method created to process image data in the form of multiple arrays [58, 59]. Convolution seeks translation equivalence and discovers feature representation using common kernel weights for each channel. The kernel of convolution is channel-specific and spatially agnostic but neglects relevance representation on the channel domain. Therefore, it cannot mine the connection among channels. Also, in case of location-related challenges, the convolution receptive area makes it difficult to record extended spatial linkages. In contrast, several DNNs have demonstrated the efficacy of kernels having cross-channel or inter-channel repetition, raising the question about the adaptability of distinct channel convolution kernels. So, it is worthwhile to examine how to efficiently exploit the various types of spatial extent and channel domain representation information. The spatial extent and channel region features of involution presented by Duo Li are the inverses of convolution, which may aid kernels in discovering the connection among the channels on the channel region to augment the fundamental feature discovering of the convolution [60]. There are currently comparatively few studies on involution. The DNNs for classical image classification continue to utilise convolution as their essential building block. Li et al. [60]

analysed the characteristics of convolution to address the issues mentioned above. A channel-independent, location-specific "involution kernel" has been proposed by the authors. Owing to the location-specific character of the kernel, they argue that self-attention falls within the involution technique by design.

The datasets we are working with contain very crucial visual features. All of the datasets are gaze or ET-based, which contains very few long-range interactions that are very important for the classification task. These small interactions are hard to detect with convolution-based networks. Involution-based models are effective when it comes to adaptively assigning the weights over different locations so as to prioritise the most impacting graphical aspects in the spatial region. We first look at the convolution process to properly derive the concept of involution. For instance, let us take a tensor X having the shape H, W, and C as the input. We gathered C convolution kernels having K, K, and C in shapes. The generated output tensor Y has the shape H, W, and C as a result of the multiply-add operation. This operation takes place between the kernels and the input tensor. The process yields an output having the forms H, W, and 3, as shown in Figure 6. The convolution kernel is location-agnostic, and it is not dependent on the spatial position of the input. Alternately, every channel in the derived output is from a unique convolution filter that makes it channel-specific. The objective is to build a channel-neutral and location-specific operation. It is challenging to execute these specific features.

We have created every kernel based on defined spatial locations to address the mentioned problem. This method assists in the processing of input tensors with varied resolutions. This process of kernel formation is shown in Figure 7. Here, C denotes the total number of channel groups, where $K \times K \times C$ filters are created. Despite applying only one filter and passing it to all C input channels, C number of filters are generated and sent to input channels. There are a total of three involution layers in our developed model. The design consists of two involution blocks, each of which has a dropout layer and a max pooling of (2 × 2). One involution layer with ReLU as the activation function comprises the first involution block. Two involution layers have the same activation function in the second involution block. Next, we flatten the 2-dimensional outputs to 1-dimensional values. The fully connected layer has 4 dense layers with 256, 96, 64, and 32 nodes. This design of the model was inspired by the architecture of the VGG-16 model [61], which has a simpler architectural design. Since the image size is small and we do not require heavyweight models and training, we follow this architecture for faster training speed (complex architecture will take more time to train). Also, as the model has a comparatively small number of layers and parameters and has a simpler architecture, it has a reduced computational complexity. For these reasons, the model is regarded as a lightweight model. It allowed the model to achieve good results using a limited computational resource. Figure 8 illustrates the architecture of the proposed model. The parameters in each layer and shape of the layers are provided in Table 1. Figures 9 and 10 show the feature maps generated on some sample scanpaths and saliency maps, respectively. In each involution layer, the ReLU activation function was paired with the Categorical cross-entropy loss function, and the Adam optimiser was utilised for optimisation.

3 | IMPLEMENTATION/ EXPERIMENTAL EVALUATION

This section is concerned with assessing the experimental results obtained by the proposed model. This segment also addresses the output comparisons on the two datasets we have used. And lastly, the uncertainty analysis of the proposed model using Monte Carlo Dropout (MCD) is investigated.

3.1 | Environment

Our proposed INN architecture and mentioned processes are developed using the Python libraries TensorFlow [62], Keras [57], Matplotlib [63], and OpenCV [64]. The model has been trained and evaluated on a GeForce RTX 3070Ti with a performance of 33.2 TeraFLOPS.

3.2 | Experiment

One of the goals of our proposed design is to generate the required outcome with the fewest number of parameters possible by keeping the size of the input form to a minimum.

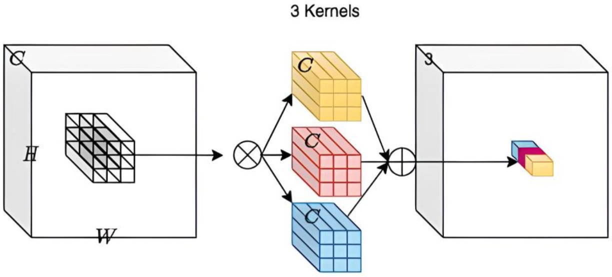

FIGURE 6 Feature extraction from an RGB image using the convolution process. It shows how convolution works by splitting the image into several parts.

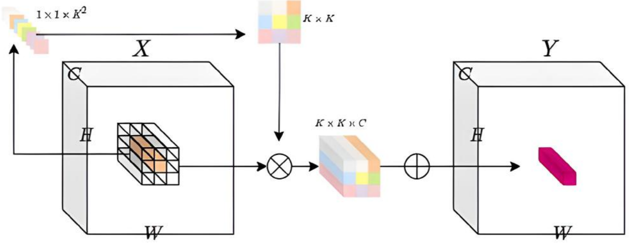

FIGURE 7 Feature extraction from an RGB image using the involution process. It shows how involution works without splitting the image into several parts.

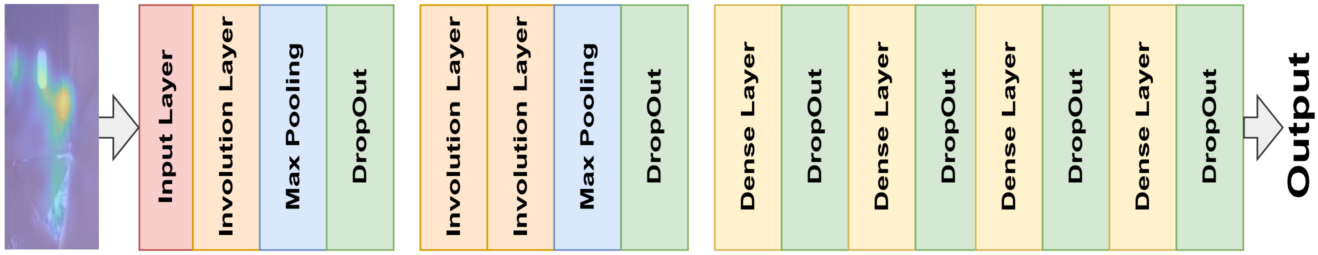

FIGURE 8 Architecture of the proposed involutional neural network method. The model comprises the Involution layer, Max pooling layer, Dropout, and Dense layer.

TABLE 1 Parameters and shape of layers of the proposed model.

Layer	Output shape	Parameter
InputLayer	[(None, 32,32,3)]	0
Involution	((None, 32,32,3), (32, 32, 9, 1, 1))	26
ReLU	(None, 32,32,3)	0
Max Pooling 2D	(None, 32, 32, 3)	0
Dropout	(None, 32, 32, 3)	0
Involution	((None, 32,32,3), (32, 32, 9, 1, 1))	26
ReLU	(None, 32,32,3)	0
Involution	((None, 32,32,3), (32, 32, 9, 1, 1))	26
ReLU	(None, 28,28,3)	0
Max Pooling 2D	(None, 32, 32, 3)	0
Monte Carlo Dropout	(None, 32, 32, 3)	0
Flatten	(None, 3072)	0
Dense	(None, 256)	786,688
Dropout	(None, 256)	0
Dense	(None, 96)	246,772
Dropout	(None, 96)	0
Dense	(None, 64)	6208
Dropout	(None, 64)	0
Dense	(None, 32)	2080
Monte Carlo Dropout	(None, 32)	0
Dense	(None, 2)	66
	Total parameters	819,792
	Trainable parameters	819,786
	Non-trainable parameters	6

A system with fewer parameters delivers improved accuracy and processing speed. The model starts with a 32 × 32 pixel picture, including three extracted channels. After that, three INN layers having kernel sizes of (3,3) were employed. In addition, Max Pooling layers having a (2,2) pool size are used. In addition, all INN layers consist of strides of 1 by default. Instead of using Tanh/Sigmoid activation functions, we adopted the ReLU activation function.

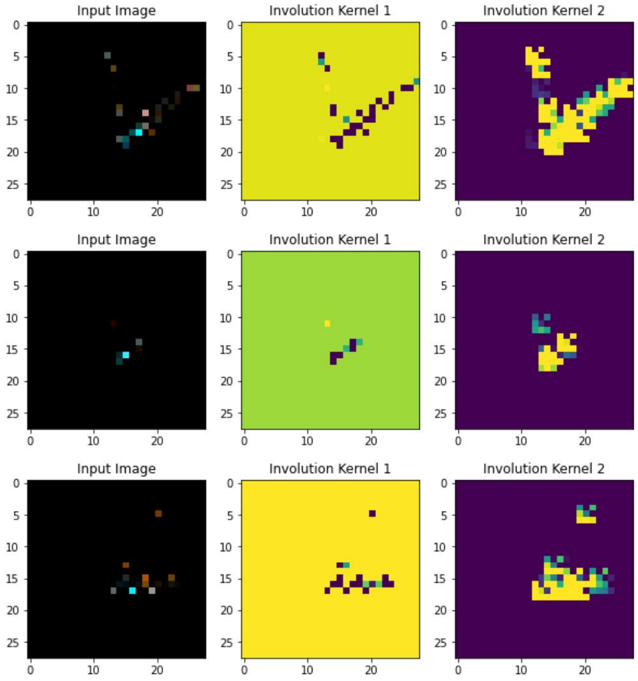

FIGURE 9 Feature Map of the model on Dataset-1. The figure presents the feature mining procedure of the involution kernel of the model from SP images.

After the transformation of the tensor into a flattened one-dimensional tensor, the first 64 nodes of the INN are replaced by fully connected FC layers. Dropouts are used to prevent overfitting in the model [65], and Adam is used as the optimiser function. The learning rate has been set to 0.0001. We have experimented using different numbers of epochs. The early stopping value has been set to 5 epochs. For Dataset-1, the model achieved optimal accuracy within only 10 epochs. For Dataset-2, 35 epochs have been required to obtain optimal accuracy. The batch size has been set to 64. The proposed design is expected to perform better since the design has fewer parameters and hence fewer computation requirements. Table 2 presents the optimised hyper-parameter values for this experiment.

3.3 | Result analysis

For result analysis, a variety of performance metrics have been utilised, including precision, recall, F1-score, support, accuracy, and ROC AUC score. The formulas for determining these metrics are shown in Equations (1)–(4), respectively.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1\ Score = \frac{2 * (Precision * Recall)}{Precision + Recall} \tag{3}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} * 100\ \% \tag{4}$$

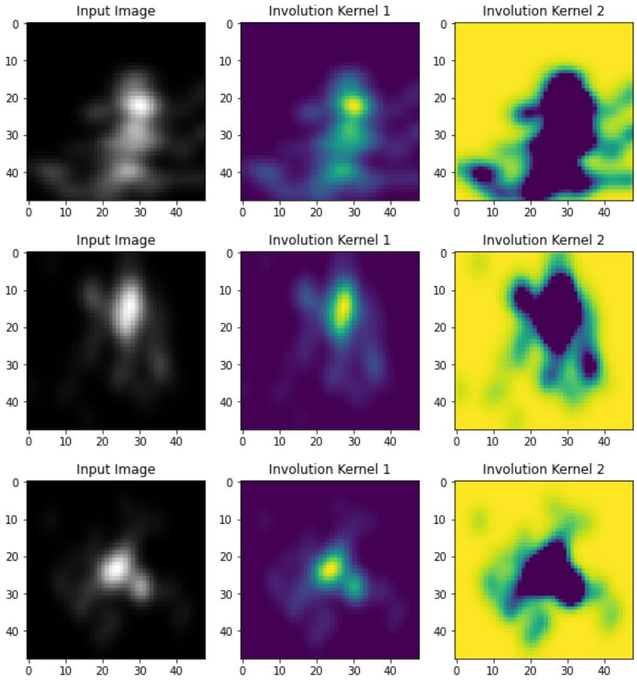

FIGURE 10 Feature Map of the model on Dataset-2 [FM]. The figure presents the feature mining procedure of the involution kernel of the model from FM images.

TABLE 2 Hyper-parameter values for training the model.

Hyper-parameter	Value
Activation function	ReLU
Initial learning rate	0.0001
Optimiser	Adam
No of epochs	35
Early stopping	5
Dropout rate	0.3–0.5
Train-test split	90%–10%
Batch size	64

In the following equations, TP stands for true positive, TF presents true negative, FP represents false positive, and FN stands for false negative predictions. Table 3 displays the precision, recall, F1-score, support, and accuracy found in the experiment.

For training and testing the model, we have used the three different types of ET images found on the dataset. In Dataset-1, there was only one type of ET image (SP). In Dataset-2, two different types of ET images (FM and HM) are available. We have conducted experiments using one type of image at a time, for example, using SP images for training and testing. The model achieved 98.12% accuracy, 97.89% precision, 99.03% recall, 98.55% F1 score, and 99.72% AUC while experimenting using Dataset-1 images. The model achieved 96.83% accuracy, 95.43% precision, 96.13% recall, 95.77% F1 score and 99.10% AUC while experimenting using Dataset-2 [FM] images. The model achieved 97.61% accuracy, 95.17% precision, 98.33% recall, 95.11% F1 score and 96.29% AUC while experimenting using Dataset-2 [HM] images.

The graph of achieved accuracy and corresponding loss for Dataset-1, Dataset-2 [FM], and Dataset-2 [HM] are shown in Figure 11. We can see from Figure 11 that the model performed well and obtained excellent accuracy in very few epochs. In the case of Dataset-1 [SP], we can see from Figure 11a that the model achieved 98.12% accuracy in only around 9–10 epochs. After 10 epochs, the accuracy has not improved much and thus the training process was stopped. Validation accuracy proves that the model has not been overfitted. In the case of Dataset-2 [FM] and Dataset-2 [HM], we can see from Figures 11b,c that the training process has become saturated in around 35 epochs while achieving 96.83% and 97.13% accuracy, respectively. From the validation curve, we can be determined that there is no or very little amount of overfitting. The training has been stopped using the early stopping method having a value of 5. The method stopped the training process when there was no improvement in accuracy in 5 epochs.

3.4 | Impact of data augmentation

Using a sufficiently large and varied dataset is one of the essential requirements for constructing a viable model. Due to its unavailability, it is not always possible to acquire a large dataset. The process of creating new and distinct data by minimally changing current data is known as data augmentation. By this method, model resilience and performance may be enhanced, while overfitting can be avoided [66]. As noted

TABLE 3 Results obtained with proposed involutional neural network model.

Dataset	Accuracy	Precision	Recall	F1 score	AUC
1 [SP]	98.12	97.89	99.03	98.55	99.72
2 [FM]	96.83	95.43	96.13	95.77	99.10
2 [HM]	97.61	95.17	98.33	95.11	96.29

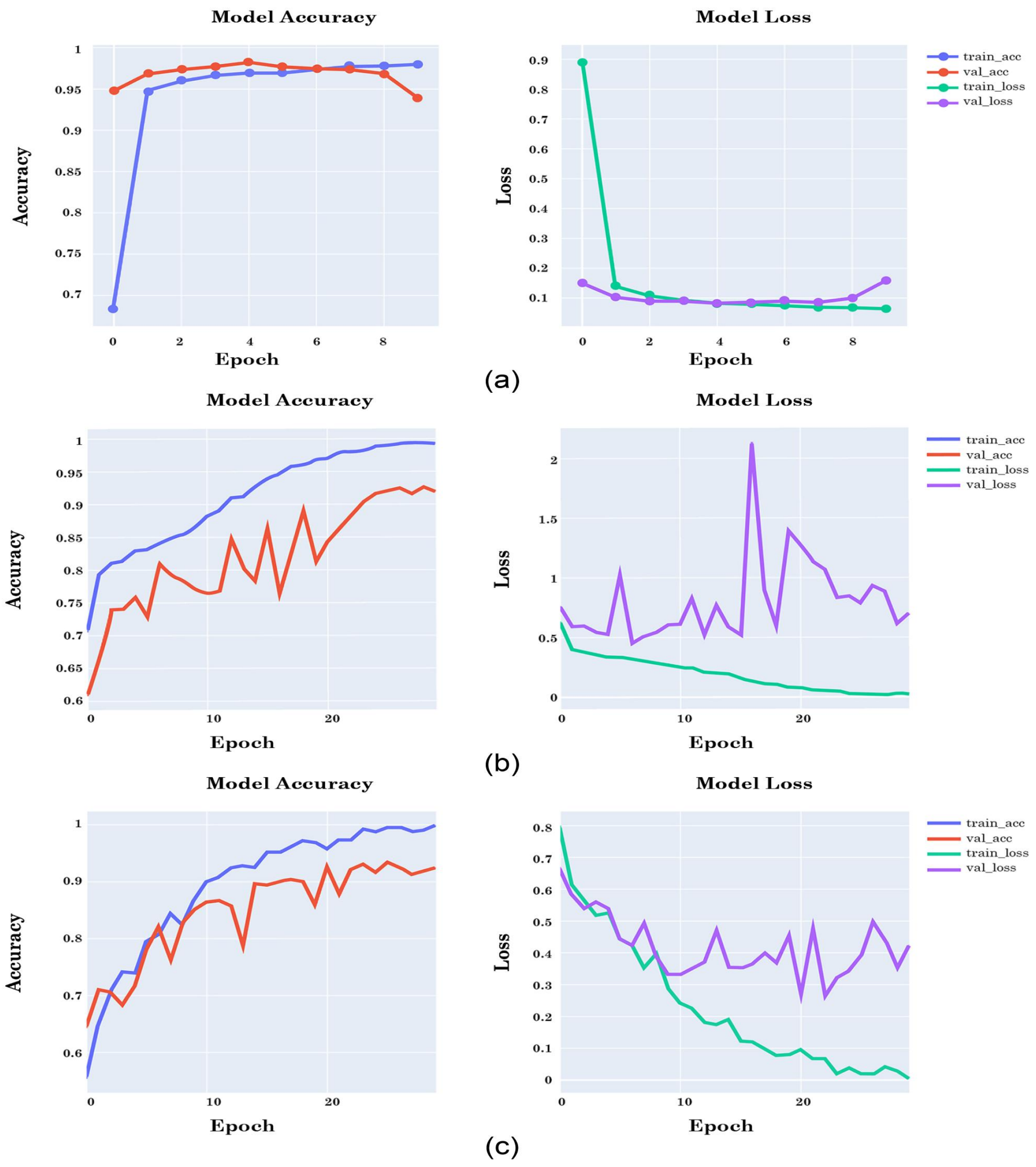

FIGURE 11 The figure presents a graph of Accuracy and Loss per epoch of three different experiments. (a) Presents graph of accuracy (left) and loss (right) over epoch on Dataset-1. (b) Presents graph of accuracy (left) and loss (right) over epoch on Dataset-2 [FM]. (c) Presents graph of accuracy (left) and loss (right) over epoch on Dataset-2 [HM].

previously, four forms of augmentation, namely rotation, flipping, scaling, and increasing brightness were done to each dataset in order to create a robust model. Table 4 presents the impact of augmentation on the datasets. From the table, we can see that the augmentation process has a huge impact on the model. After the implementation of augmentation, the model accuracy improved by around 10% in each case. Augmentation has also helped the model to achieve more generalisation.

TABLE 4 Results obtained with proposed method (impact of image augmentation).

Dataset	Before augmentation	After augmentation
Dataset-1 [SP]	95.25	98.12
Dataset-2 [FM]	93.83	96.83
Dataset-2 [HM]	95.42	97.61

3.5 | Comparison with state-of-the-art image classification models

The proposed model is compared with other existing SOTA image classification models with respect to different evaluation metrics. The models we have used for comparison are Vision Transformer [67], Swin Transformer [68], Compact Convolutional Transformer [69], ConvMixer [70], InceptionV3 [71], VGG16 [61], VGG19 [61], ResNet50 [72], EfficeintNetB7 [73], MobileNetV2 [74], Xception [75], and DenseNet201 [76]. Table 5 presents the results obtained by the SOTA models for image classification. The proposed INN model outperforms

TABLE 5 Performance of SOTA image classification models on Dataset-1 and Dataset-2.

Dataset	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Dataset 1	**Proposed**	**98.12**	**97.89**	**99.03**	**98.55**
	ViT	92	91.75	93.75	92.75
	SWT	93.8	92.75	92.75	92.75
	CCT	96.74	94.75	94.75	94.75
	InceptionV3	68.68	72.56	64.45	68.26
	VGG16	83	80.89	84.04	82.43
	VGG19	65.84	66.64	66.64	82.43
	ResNet50	94.25	94.12	94.12	94.12
	EfficientNetB7	59.77	59.74	59.74	59.74
	MobileNetV2	59.77	59.9	59.9	59.9
	Xception	82.18	82.5	82.18	82.33
	DenseNet201	91.95	91.96	92.21	92.08
Dataset 2 [fixation maps]	**Proposed**	**96.83**	**95.43**	**96.13**	**95.77**
	ViT	86.12	87.65	88.77	88.2
	SWT	87.55	88.33	88.33	88.33
	CCT	90.23	89.63	89.63	89.63
	InceptionV3	68.68	72.56	64.45	68.26
	VGG16	87.22	81.55	85.55	83.55
	VGG19	65.84	66.64	66.64	82.43
	ResNet50	93.8	94.2	94.8	94.5
	EfficientNetB7	55.36	58.18	58.18	58.18
	MobileNetV2	56.77	55.7	55.9	55.8
	Xception	80.18	80.79	79.38	80.07
	DenseNet201	90.95	92.76	91.41	92.08
Dataset 2 [heatmaps]	**Proposed**	**97.61**	**95.17**	**98.33**	**95.11**
	ViT	85.76	87.13	85.17	87.3
	SWT	87.55	88.33	88.33	87.33
	CCT	88.56	89.55	89.76	87.63
	InceptionV3	71.66	73.56	70.45	69.26
	VGG16	88.22	89.34	85.55	87.55
	VGG19	69.84	67.64	68.74	70.43
	ResNet50	92.8	92.2	92.8	91.5
	EfficientNetB7	58.56	57.18	59.18	58.48
	MobileNetV2	62.71	61.71	62.92	62.28
	Xception	79.22	80.43	79.18	79.07
	DenseNet201	90.45	90.16	91.41	90.08

Note: The bold line indicates the highest value.

the existing SOTA image classification models in every case, which can be seen from Table 5. The bold line indicates the highest value. From Figure 12, we can see that the proposed models consists of much less number of parameters than the other models. Thus, the proposed model took less time for computation in training and testing.

3.6 | Comparison with existing models

The proposed INN model has been compared with existing models suggested in the detection of ASD with respect to different evaluation metrics. We have compared our work with other existing literature where either ASD Dataset-1 [53] or ASD Dataset-2 [54] has been used. Thus, the INN proposed model has been compared with the models that have been implemented on Dataset-1 and Dataset-2 separately. Table 6 presents the comparison among models implemented on Dataset-1, and Table 7 presents a comparison among the models implemented on Dataset-2. As the existing literature utilised FM images for the experiment, the comparison has been done on FM images only. We have used accuracy and AUC as the evaluation metrics for comparison. From the comparisons, it has been found that being a lightweight model, the proposed model outperforms the existing models.

3.7 | Uncertainty analysis

Often, the dropout strategy is employed to minimise model complexity and prevent overfitting [65]. During training, a dropout layer divides the results provided by a node by a Bernoulli-distributed binary mask, randomly setting the number of neurons in the DNN to zero. The trained neural network that was not dropped is then applied during test time. Gal and Ghahramani established that dropout during testing is an approximation of probabilistic Bayesian models in deep Gaussian processes [92]. MCD measures the deviation of

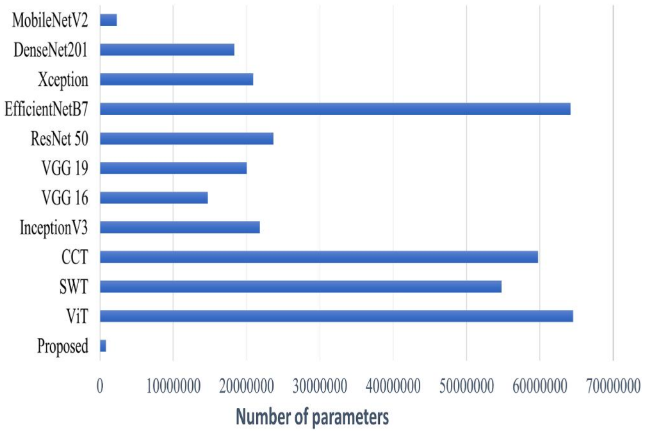

FIGURE 12 Comparison among various SOTA image classification models in terms of the number of parameters.

TABLE 6 Comparison among different works on Dataset 1.

Reference	Model	Accuracy	AUC
Proposed model	Deep INN	98.12	99.72
Carette et al. [53]	Logistic regression	-	81.95
Carette et al. [35]	Single layer artificial neural network	-	92.01
Elbattah et al. [77]	k-means clustering	94.00	-
Elbattah et al. [42]	Variational autoencoders	70.01	76.10
Akter et al. [33]	Clustering and multilayer perceptron	87.00	79.00
Xie et al. [41]	Two stream deep learning network	95.01	-
Gaspar et al. [78]	Kernel extreme learning machine	98.80	-
Cilia et al. [31]	CNN	71.04	-
Mumenin et al. [79]	CNN	97.41	99.60
Kanhirakadavath and Chandran [40]	CNN + DNN	-	97.00

TABLE 7 Comparison among different works using Dataset-2.

Author	Model information	Accuracy	AUC
Proposed model	Deep INN	96.83	99.10
Mazumdar et al. [30]	TreeBagger classifier	68.50	-
Shi Chen and Qi Zhao [80]	CNN + LSTM	93.00	98.00
Liaqat et al. [81]	Modified ResNet50	62.13	64.4
Tao and Shyu [82]	CNN-LSTM	57.90	56.97
Fang et al. [83]	Dilated CNN + LSTM	79.94	79.00
Arru et al. [84]	Decision Tree (TreeBagger)	59.30	59.50
Startsev and Dorr [85]	Random forest	63.90	-
Tamilarasi and Shanmugam [86]	CNN	89.20	-
Xie et al. [87]	Two stream CNN (VGG-16)	95.00	-
Wu et al. [88]	Deep CNN(ResNet)	65.41	-
Wei et al. [44]	Dynamic Filter + LSTM	61.48	64.33
Hriti et al. [89]	CNN + ANN	93.00	93.00
Shihab et al. [34]	PCA	95.10	-
Mahalakshmi and Praveena [43]	CNN	75.23	-
Kang et al. [90]	Support vector machine	85.00	-
Alsaade and Alzahrani [46]	Transfer Learning (Xception)	91.11	-
Ahmed et al. [91]	CNN	95.54	-

predictions of the model from their forecast distribution by sampling T new dropout masks for every forward pass. The set may then be viewed as samples obtained from the output distribution, which is valuable for extracting information about the variability of the output. This knowledge is useful in decision-making. In reality, characterising the model's uncertainty may permit distinct treatment of uncertain inputs. The computational complexity of MCD is proportional to the total number of forward passes T, which is its primary downside. Alternately, the forward passes might be executed simultaneously which results in a constant running time. In addition, if the MCD layers are positioned close to the output layer, the input of the first dropout layer may be preserved in the first pass so that it can be reused in subsequent runs, therefore reducing unnecessary computation. Hence, the complexity may be drastically lowered, allowing it to be used in real-time applications. The MCD model estimate is equal to the mean of T forecasts.

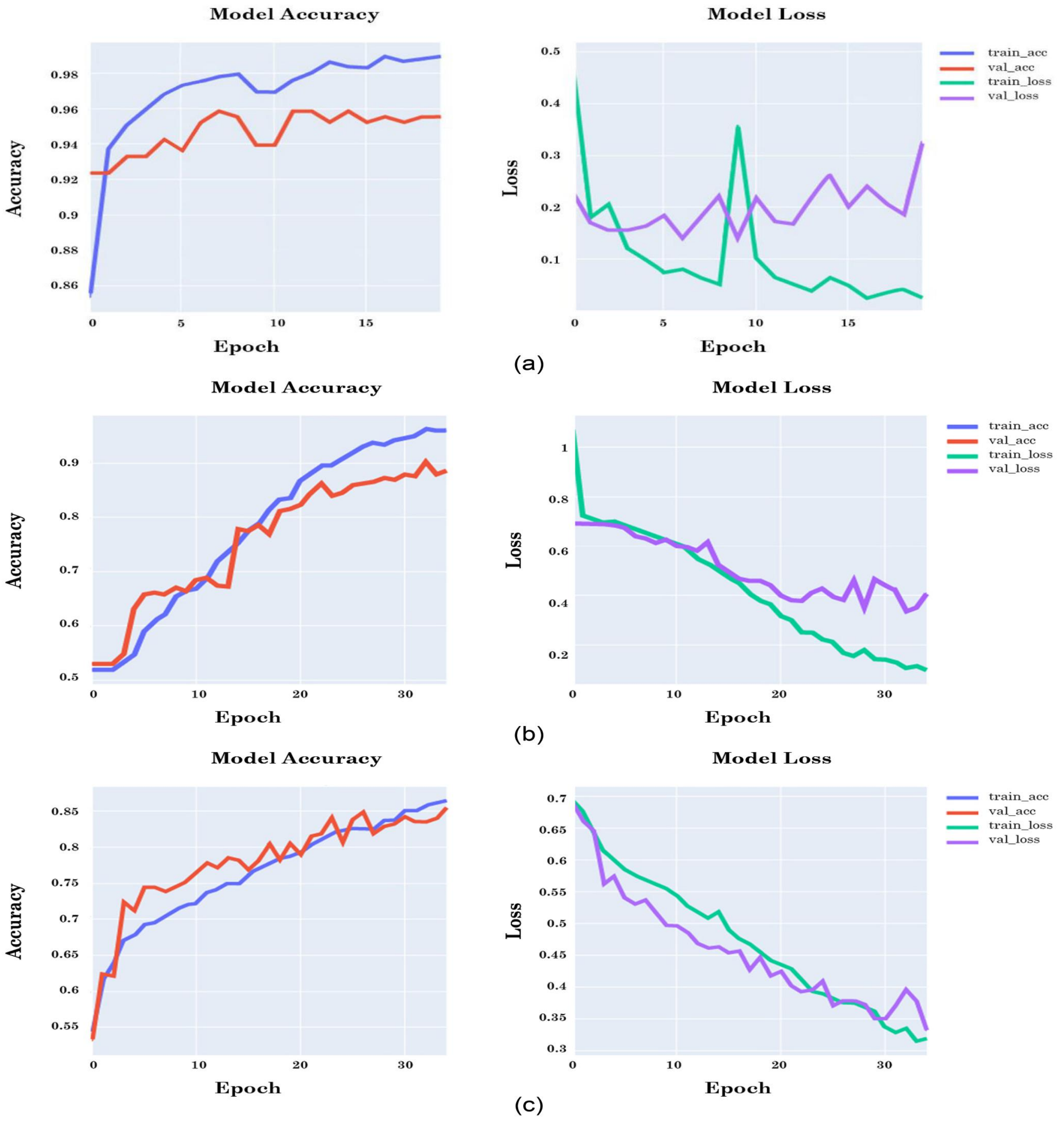

FIGURE 13 Graph for (a) accuracy and (b) loss over epochs after using Monte Carlo dropout on Dataset-1 [SP].

$$\mathbf{p}^* = \frac{1}{T}\sum_{t=1}^{T}\mathbf{p}_t \tag{5}$$

According to the authors in Ref. [92], $T = 50$ is a safe option for estimating uncertainty; however, this number must also be examined in light of MCD's prediction ability. The MCD may be seen as a specific instance of Deep Ensembles (training several comparable networks and sampling outputs from every one of them), which is an additional method to enhance the performance of DL models and quantify uncertainty. To calculate the distribution of outputs using MCD, we embedded models with a dropout rate of 50%. We utilised 200 test samples and predict each sample 400 times (MC Sampling). This is necessary for determining the uncertainty associated with the predicted class-wise score distribution of 200 test samples. This ensemble accuracy, unlike the standard accuracy score, is determined by Monte Carlo sampling with 500 sample data.

In the case of Dataset-1, the model achieved 98.28% accuracy. After performing MCD, the accuracy improved to 98.33%. The accuracy and loss graph for Dataset-1 has been presented in Figure 13a. In the case of Dataset-2 [FM] and Dataset-2 [HM], the accuracy has improved to 97.11% and 97.88%, respectively. The accuracy and loss graph for Dataset-2 [FM] and Dataset-2 [HM] has been presented in Figures 13b,c, which also shows the improvement in accuracy and loss.

From the Monte Carlo-Ensemble accuracies and plots, we can see that the model is most of the time giving more than 98% for Dataset-1, 92% for Dataset-2(FM), and 81% for Dataset-2(HM). Figures 14a–c show the distribution of the Monte Carlo predictions (blue) and prediction of the ensemble (red) for Dataset-1, Dataset-2 [FM], and Dataset-2 [HM], respectively. This suggests that our proposed INN model is reliable.

4 | DISCUSSION

In this work, we have proposed an INN-based lightweight robust DL model for ASD diagnosis. The INN model has been trained and evaluated using four different types of ET images from two different datasets (ASD Dataset-1 [53] and ASD Dataset-2 [54]). To ensure the effectiveness of the model, it has been evaluated using several evaluation metrics. From Table 3, we can see that the model performed well in the case of all four types of images. In Dataset-1, the proposed model achieved 98.12% accuracy using scanpath images. And in Dataset-2, the model achieved 96.83%, 97.61%, and 95.23% accuracy while using a fixation map, heatmap and fixation point, respectively. It ensures the generalisation capability and robustness of the model. After that, we investigated the impact of data augmentation. It is noticeable that after augmentation, the performance of the model increased significantly. Then, the model has been compared with the SOTA image classification models. The proposed model outperformed the SOTA image classification models by almost a 2% margin in the case of accuracy. Brief experimental results have been given in Table 5. After that, the experimental results of the proposed INN model have been compared with the models from the existing literature. Table 6 presents the comparison among the works conducted on Dataset-1 [53], and Table 6 presents the comparison among the works conducted on Dataset-2 [93].

5 | CONCLUSION

This research presents an automated categorisation approach for early and accurate diagnosis in order to prevent the horrifying consequences of ASD. Involutional Neural Network has been used as the main block for feature extraction from images. The main reason behind using INN is its channel-independent and location-specific capability, which is able to extract features more

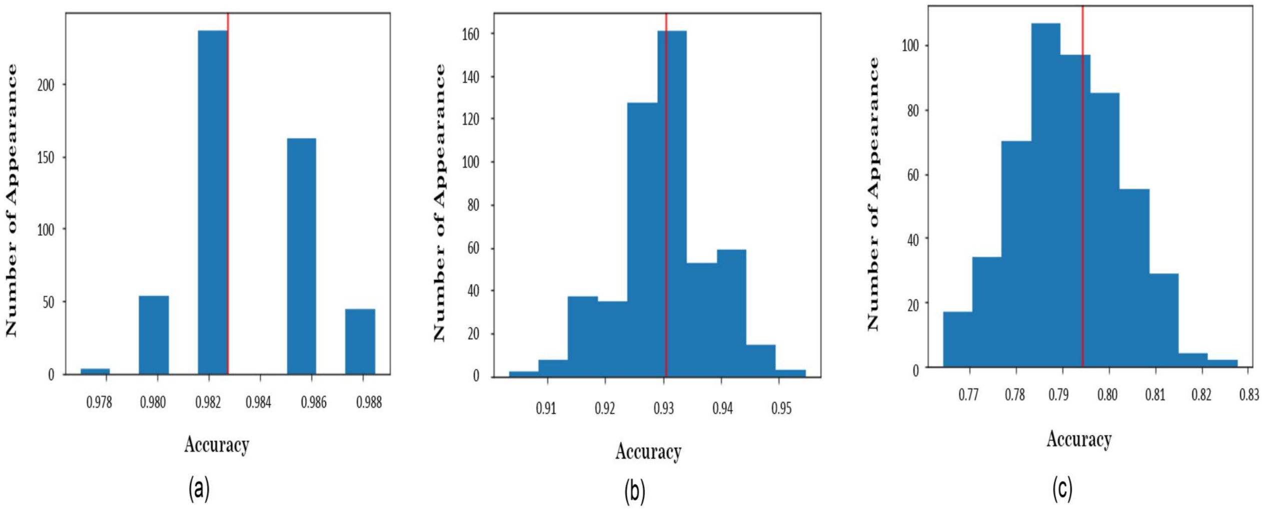

FIGURE 14 Distribution of the Monte Carlo prediction (Blue) and ensemble prediction (Red) on (a) Dataset-1, (b) Dataset-2 [FM] and (c) Dataset-2 [HM].

efficiently. Two publicly available datasets containing four types of images, namely ET scanpath, fixation maps, and heatmaps, have been used to train and evaluate the model. We have conducted experiments and evaluated the model on various evaluation metrics. The proposed INN model has outperformed the existing SOTA image classification models and other existing research conducted on the used datasets. In Dataset-1, the proposed model achieved 98.12% accuracy. In Dataset-2, it achieved 96.83%, and 97.61% accuracy using fixation map and heatmap, respectively. The model has a comparatively low number of parameters compared to SOTA models and a reduced computational complexity due to its simple architecture. Uncertainty estimation using MCD ensures the reliability of the prediction made by the model.

AUTHOR CONTRIBUTIONS

Nasirul Mumenin: Conceptualisation; data curation; formal analysis; investigation; methodology; resources; writing—original draft; writing—review and editing. **Mohammad Abu Yousuf**: Project administration; supervision; validation; writing—original draft; writing—review and editing. **Md Asif Nashiry**: Supervision; validation; writing—original draft; writing—review and editing. **AKM Azad**: Validation; writing—review and editing. **Salem A. Alyami**: Validation; writing—review and editing. **Pietro Lio'**: Validation; writing—review and editing. **Mohammad Ali Moni**: Supervision; validation; writing—review and editing.

ACKNOWLEDGEMENTS

This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-RP23004).

CONFLICT OF INTEREST STATEMENT

There is no conflict of interest among the authors.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Dataset-1: [fighsare data repository [at [https://www.scitepress.org/Link.aspx?doi&equals;10.5220/0007402601030112], reference number [53] and Dataset-2: [fighsare data repository [at [https://doi.org/10.5281/zenodo.2647418], reference number [54].

ORCID

Nasirul Mumenin https://orcid.org/0000-0002-8615-8348
Mohammad Ali Moni https://orcid.org/0000-0003-0756-1006

REFERENCES

1. Lord, C., et al.: Autism spectrum disorder. Nat. Rev. Dis. Prim. 6(1), 1–23 (2020). https://doi.org/10.1038/s41572-019-0138-4
2. Christensen, D.L., et al.: Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2012. MMWR Surveill. Summ. 65(13), 1–23 (2018). https://doi.org/10.15585/mmwr.ss6503a1
3. Autism. https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders
4. Maenner, M.J., et al.: Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2018. MMWR Surveill. Summ. 70(11), 1–16 (2021). https://doi.org/10.15585/mmwr.ss7011a1
5. Psychiatry.org - what Is Autism Spectrum Disorder? Available from:. https://psychiatry.org:443/patients-families/autism/what-is-autism-spectrum-disorder
6. Koegel, L.K., et al.: The importance of early identification and intervention for children with or at risk for autism spectrum disorders. Int. J. Speech Lang. Pathol. 16(1), 50–56 (2014). https://doi.org/10.3109/17549507.2013.861511
7. Charman, T., Baird, G.: Practitioner review: diagnosis of autism spectrum disorder in 2-and 3-year-old children. JCPP (J. Child Psychol. Psychiatry) 43(3), 289–305 (2002). https://doi.org/10.1111/1469-7610.00022
8. Constantino, J.N., Charman, T.: Diagnosis of autism spectrum disorder: reconciling the syndrome, its diverse origins, and variation in expression. Lancet Neurol. 15(3), 279–291 (2016). https://doi.org/10.1016/s1474-4422(15)00151-9
9. Zwaigenbaum, L., Penner, M.: Autism spectrum disorder: advances in diagnosis and evaluation. Br. Med. J., 361 (2018)
10. Akshoomoff, N., Corsello, C., Schmidt, H.: The role of the autism diagnostic observation schedule in the assessment of autism spectrum disorders in school and community settings. Calif. Sch. Psychol. 11(1), 7–19 (2006). https://doi.org/10.1007/bf03341111
11. Eslami, T., et al.: Asd-diagnet: a hybrid learning approach for detection of autism spectrum disorder using fmri data. Front. Neuroinf. 13, 70 (2019). https://doi.org/10.3389/fninf.2019.00070
12. Philip, R.C., et al.: A systematic review and meta-analysis of the fmri investigation of autism spectrum disorders. Neurosci. Biobehav. Rev. 36(2), 901–942 (2012). https://doi.org/10.1016/j.neubiorev.2011.10.008
13. Bosl, W., et al.: Eeg complexity as a biomarker for autism spectrum disorder risk. BMC Med. 9(1), 1–16 (2011)
14. Bosl, W.J., Tager Flusberg, H., Nelson, C.A.: Eeg analytics for early detection of autism spectrum disorder: a data-driven approach. Sci. Rep. 8(1), 1–20 (2018)
15. Zecavati, N., Spence, S.J.: Neurometabolic disorders and dysfunction in autism spectrum disorders. Curr. Neurol. Neurosci. Rep. 9(2), 129–136 (2009). https://doi.org/10.1007/s11910-009-0021-x
16. Frye, R.E., et al.: Biomarkers of abnormal energy metabolism in children with autism spectrum disorder. N. Am. J. Med. Sci. 5(3), 141 (2012). https://doi.org/10.7156/v5i3p141
17. Loth, E., et al.: Facial expression recognition as a candidate marker for autism spectrum disorder: how frequent and severe are deficits? Mol. Autism. 9(1), 1–11 (2018). https://doi.org/10.1186/s13229-018-0187-7
18. Wallace, S., Coleman, M., Bailey, A.: An investigation of basic facial expression recognition in autism spectrum disorders. Cognit. Emot. 22(7), 1353–1380 (2008). https://doi.org/10.1080/02699930701782153
19. Norbury, C.F., et al.: Eye-movement patterns are associated with communicative competence in autistic spectrum disorders. JCPP (J. Child Psychol. Psychiatry) 50(7), 834–842 (2009). https://doi.org/10.1111/j.1469-7610.2009.02073.x
20. Wan, G., et al.: Applying eye tracking to identify autism spectrum disorder in children. J. Autism Dev. Disord. 49(1), 209–215 (2019). https://doi.org/10.1007/s10803-018-3690-y
21. Papagiannopoulou, E.A., et al.: A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders. Soc. Neurosci. 9(6), 610–632 (2014). https://doi.org/10.1080/17470919.2014.934966
22. Yarbus, A.L.: Eye Movements and Vision. Springer (2013)
23. Zammarchi, G., Conversano, C.: Application of eye tracking technology in medicine: a bibliometric analysis. Vision 5(4), 56 (2021). https://doi.org/10.3390/vision5040056
24. Posner, M.I.: Orienting of attention. Q. J. Exp. Psychol. 32(1), 3–25 (1980). https://doi.org/10.1080/00335558008248231

25. Wadhera, T., Kakkar, D.: Eye tracker: an assistive tool in diagnosis of autism spectrum disorder. In: Emerging Trends in the Diagnosis and Intervention of Neurodevelopmental Disorders, pp. 125–152. IGI Global (2019)
26. Frith, U., Happé, F.: Autism spectrum disorder. Curr. Biol. 15(19), R786–R790 (2005). https://doi.org/10.1016/j.cub.2005.09.033
27. Bataineh, E., et al.: Visual attention toward socially rich context information for autism spectrum disorder (asd) and normal developing children: an eye tracking study. In: Proceedings of the 16th International Conference on Advances in Mobile Computing and Multimedia, pp. 61–65 (2018)
28. Solovyova, A., et al.: Early Autism Spectrum Disorders Diagnosis Using Eye-Tracking Technology. *arXiv preprint arXiv:200809670*, (2020)
29. Eraslan, S., et al.: Autism detection based on eye movement sequences on the web: a scanpath trend analysis approach. In: Proceedings of the 17th International Web for All Conference, pp. 1–10 (2020)
30. Mazumdar, P., Arru, G., Battisti, F.: Early detection of children with autism spectrum disorder based on visual exploration of images. Signal Process. Image Commun. 94, 116184 (2021). https://doi.org/10.1016/j.image.2021.116184
31. Cilia, F., et al.: Computer-aided screening of autism spectrum disorder: eye-tracking study using data visualization and deep learning. JMIR Hum. Factors 8(4), e27706 (2021). https://doi.org/10.2196/27706
32. Almourad, M.B., Bataineh, E.: Visual attention toward human face recognizing for autism spectrum disorder and normal developing children: an eye tracking study. In: Proceedings of the 2020 the 6th International Conference on E-Business and Applications, pp. 99–104 (2020)
33. Akter, T., et al.: Machine learning model to predict autism investigating eye-tracking dataset. In: 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), pp. 383–387. IEEE (2021)
34. Shihab, A.I., Dawood, F.A., Kashmar, A.H.: Data analysis and classification of autism spectrum disorder using principal component analysis. Adv. Bioinformatics, 2020 (2020)
35. Carette, R., et al.: Learning to predict autism spectrum disorder based on the visual patterns of eye-tracking scanpaths. In: HEALTHINF, pp. 103–112 (2019)
36. Carette, R., et al.: Automatic autism spectrum disorder detection thanks to eye-tracking and neural network-based approach. In: International Conference on IoT Technologies for Healthcare, pp. 75–81. Springer (2017)
37. Akter, T., et al.: Machine learning-based models for early stage detection of autism spectrum disorders. IEEE Access 7, 166509–166527 (2019). https://doi.org/10.1109/access.2019.2952609
38. Uddin, M.J., et al.: An integrated statistical and clinically applicable machine learning framework for the detection of autism spectrum disorder. Computers 12(5), 92 (2023). https://doi.org/10.3390/computers12050092
39. Lin, P.I., et al.: Identifying subgroups of patients with autism by gene expression profiles using machine learning algorithms. Front. Psychiatr. 12, 637022 (2021). https://doi.org/10.3389/fpsyt.2021.637022
40. Kanhirakadavath, M.R., Chandran, M.S.M.: Investigation of eye-tracking scan path as a biomarker for autism screening using machine learning algorithms. Diagnostics 12(2), 518 (2022). https://doi.org/10.3390/diagnostics12020518
41. Xie, J., et al.: A Two-Stream End-To-End Deep Learning Network for Recognizing Atypical Visual Attention in Autism Spectrum Disorder. *arXiv preprint arXiv:191111393*, (2019)
42. Elbattah, M., et al.: Nlp-based approach to detect autism spectrum disorder in saccadic eye movement. In: 2020 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1581–1587. IEEE (2020)
43. Praveena, K., Mahalakshmi, R.: Classification of autism spectrum disorder and typically developed children for eye gaze image dataset using convolutional neural network. Int. J. Adv. Comput. Sci. Appl. 13(3) (2022). https://doi.org/10.14569/ijacsa.2022.0130345
44. Wei, W., et al.: Identify autism spectrum disorder via dynamic filter and deep spatiotemporal feature extraction. Signal Process. Image Commun. 94, 116195 (2021). https://doi.org/10.1016/j.image.2021.116195
45. Duan, H., et al.: Visual attention analysis and prediction on human faces for children with autism spectrum disorder. ACM Trans. Multimed Comput. Commun. Appl. 15(3s), 1–23 (2019). https://doi.org/10.1145/3337066
46. Alsaade, F.W., Alzahrani, M.S.: Classification and detection of autism spectrum disorder based on deep learning algorithms. Comput. Intell. Neurosci., 2022 (2022)
47. Guillon, Q., et al.: Visual social attention in autism spectrum disorder: insights from eye tracking studies. Neurosci. Biobehav. Rev. 42, 279–297 (2014). https://doi.org/10.1016/j.neubiorev.2014.03.013
48. Alcañiz, M., et al.: Eye gaze as a biomarker in the recognition of autism spectrum disorder using virtual reality and machine learning: a proof of concept for diagnosis. Autism Res. 15(1), 131–145 (2022). https://doi.org/10.1002/aur.2636
49. Pierce, K., et al.: Eye tracking reveals abnormal visual preference for geometric images as an early biomarker of an autism spectrum disorder subtype associated with increased symptom severity. Biol. Psychiatr. 79(8), 657–666 (2016). https://doi.org/10.1016/j.biopsych.2015.03.032
50. Ahmed, Z.A.T., Jadhav, M.E.: A review of early detection of autism based on eye-tracking and sensing technology. In: 2020 International Conference on Inventive Computation Technologies (ICICT), pp. 160–166. IEEE (2020)
51. Vargas Cuentas, N.I., et al.: Developing an eye-tracking algorithm as a potential tool for early diagnosis of autism spectrum disorder in children. PLoS One 12(11), e0188826 (2017). https://doi.org/10.1371/journal.pone.0188826
52. Falck Ytter, T., Bölte, S., Gredebäck, G.: Eye tracking in early autism research. J. Neurodev. Disord. 5(1), 1–13 (2013)
53. Carette, R., et al.: Visualization of eye-tracking patterns in autism spectrum disorder: method and dataset. In: 2018 Thirteenth International Conference on Digital Information Management (ICDIM), pp. 248–253. IEEE (2018)
54. Duan, H., et al.: A dataset of eye movements for the children with autism spectrum disorder. In: Proceedings of the 10th ACM Multimedia Systems Conference, pp. 255–260 (2019)
55. Mele, M.L., Federici, S.: Gaze and eye-tracking solutions for psychological research. Cognit. Process. 13(S1), 261–265 (2012). https://doi.org/10.1007/s10339-012-0499-z
56. Gutiérrez, J., et al.: Saliency4asd: challenge, dataset and tools for visual attention modeling for autism spectrum disorder. Signal Process. Image Commun. 92, 116092 (2021). https://doi.org/10.1016/j.image.2020.116092
57. Gulli, A., Pal, S.: Deep Learning with Keras. Packt Publishing Ltd (2017)
58. Albawi, S., Mohammed, T.A., Al Zawi, S.: Understanding of a convolutional neural network. In: 2017 International Conference on Engineering and Technology (ICET), pp. 1–6. IEEE (2017)
59. O'Shea, K., Nash, R.: An Introduction to Convolutional Neural Networks. *arXiv preprint arXiv:151108458*. (2015)
60. Li, D., et al.: Involution: Inverting The Inherence of Convolution for Visual Recognition. arXiv. (2021). https://arxiv.org/abs/2103.06255
61. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv:14091556*, (2014)
62. Abadi, M., et al.: Tensorflow: a system for large-scale machine learning. In: Osdi, vol. 16, pp. 265–283. Savannah, GA, USA (2016)
63. Hunter, J.D.: Matplotlib: a 2d graphics environment. Comput. Sci. Eng. 9(03), 90–95 (2007). https://doi.org/10.1109/mcse.2007.55
64. Bradski, G.: The opencv library. Dr. Dobb's J. Softw. Tools Prof. Program. 25(11), 120–123 (2000)
65. Srivastava, N., et al.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15(1), 1929–1958 (2014)
66. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. J. Big Data 6(1), 1–48 (2019). https://doi.org/10.1186/s40537-019-0197-0
67. Dosovitskiy, A., et al.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv preprint arXiv:201011929*. (2020)
68. Liu, Z., et al.: Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022 (2021)

69. Hassani, A., et al.: Escaping the Big Data Paradigm with Compact Transformers. *arXiv preprint arXiv:210405704*, (2021)
70. Trockman, A., Kolter, J.Z.: Patches Are All You Need? *arXiv preprint arXiv:220109792*. (2022)
71. Szegedy, C., et al.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)
72. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
73. Tan, M., Le, Q.: Efficientnet: rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning, pp. 6105–6114. PMLR (2019)
74. Sandler, M., et al.: Mobilenetv2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520 (2018)
75. Chollet, F.: Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258 (2017)
76. Huang, G., et al.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
77. Elbattah, M., et al.: Learning clusters in autism spectrum disorder: image-based clustering of eye-tracking scanpaths with deep autoencoder. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1417–1420. IEEE (2019)
78. Gaspar, A., et al.: An optimized kernel extreme learning machine for the classification of the autism spectrum disorder by using gaze tracking images. Appl. Soft Comput. 120, 108654 (2022). https://doi.org/10.1016/j.asoc.2022.108654
79. Mumenin, N., et al.: Diagnosis of autism spectrum disorder through eye movement tracking using deep learning. In: Proceedings of International Conference on Information and Communication Technology for Development: ICICTD 2022, pp. 251–262. Springer (2023)
80. Chen, S., Zhao, Q.: Attention-based autism spectrum disorder screening with privileged modality. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1181–1190 (2019)
81. Liaqat, S., et al.: Predicting asd diagnosis in children with synthetic and image-based eye gaze data. Signal Process. Image Commun. 94, 116198 (2021). https://doi.org/10.1016/j.image.2021.116198
82. Tao, Y., Shyu, M.L.: Sp-asdnet: cnn-lstm based asd classification model using observer scanpaths. In: 2019 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pp. 641–646. IEEE (2019)
83. Fang, Y., et al.: Identifying children with autism spectrum disorder based on gaze-following. In: 2020 IEEE International Conference on Image Processing (ICIP), pp. 423–427. IEEE (2020)
84. Arru, G., Mazumdar, P., Battisti, F.: Exploiting visual behaviour for autism spectrum disorder identification. In: 2019 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pp. 637–640. IEEE (2019)
85. Startsev, M., Dorr, M.: Classifying autism spectrum disorder based on scanpaths and saliency. In: 2019 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pp. 633–636. IEEE (2019)
86. Tamilarasi, F.C., Shanmugam, J.: Convolutional neural network based autism classification. In: 2020 5th International Conference on Communication and Electronics Systems (ICCES), pp. 1208–1212. IEEE (2020)
87. Xie, J., et al.: Identifying visual attention features accurately discerning between autism and typically developing: a deep learning framework. Interdiscipl. Sci. Comput. Life Sci. 14(3), 639–651 (2022). https://doi.org/10.1007/s12539-022-00510-6
88. Wu, C., et al.: Predicting autism diagnosis using image with fixations and synthetic saccade patterns. In: 2019 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pp. 647–650. IEEE (2019)
89. Sadaf, N., et al.: Autism classification using visual and behavioral data. medRxiv, 2022 (2022)
90. Kang, J., et al.: The identification of children with autism spectrum disorder by svm approach on eeg and eye-tracking data. Comput. Biol. Med. 120, 103722 (2020). https://doi.org/10.1016/j.compbiomed.2020.103722
91. Ahmed, Z.A., Jadhav, M.E.: Convolutional neural network for prediction of autism based on eye-tracking scanpaths. Int. J. Psychosoc. Rehabil. 24(05), 2683–2689 (2020). https://doi.org/10.37200/ijpr/v24i5/pr201970
92. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: International Conference on Machine Learning, pp. 1050–1059. PMLR (2016)
93. Duan, H., et al.: Learning to predict where the children with asd look. In: 2018 25th Ieee International Conference on Image Processing (ICIP), pp. 704–708. IEEE (2018)

How to cite this article: Mumenin, N., et al.: ASDNet: a robust involution-based architecture for diagnosis of autism spectrum disorder utilising eye-tracking technology. IET Comput. Vis. 18(5), 666–681 (2024). https://doi.org/10.1049/cvi2.12271