Research Article

Journal of Current Trends in Computer Science Research

Identification of Autism Spectrum Disorder using Residual Attention Network for Facial Image Analysis.

C. Gnanaprakasam [1*], Manoj Kumar Rajagopal[2]

[a] *Assistant Professor, Dept of AI&DS, Panimalar Engineering College,Chennai,Tamil Nadu,India,*

[b]*Associate Professor, School of Electronics Engineering,Vellore Institute of Technology, Chennai, Tamil Nadu, India*

[*]Corresponding Author:
C. Gnanaprakasam, Assistant Professor, Dept of AI&DS, Panimalar Engineering College, Chennai, Tamil Nadu, India.

Submitted: 31 Jan 2023; **Accepted**: 06 Feb 2023: **Published**: 27 Feb 2023

Citation: Gnanaprakasam, C., Rajagopal, M.,K., (2023). Identification of Autism Spectrum Disorder using Residual Attention Network for Facial Image Analysis, *J Curr Trends Comp Sci Res,* 2(1), 31-39.

Abstract

The goal of the present paper is to apply deep learning algorithms to identify autism spectrum disorder (ASD) patients from a face image dataset based solely on the patient's face activation patterns. We investigated ASD Patients' face imaging data from a worldwide multispecialty database known as Autism Face Imaging Data Exchange. ASD is a brain-based disorder characterized by social deficiency and the symptoms are different scenarios. According to recent Centers for Disease Control data, ASD affects About 1 in 54 children who have been identified with autism spectrum disorder (ASD) according to estimates from CDC's Autism and Developmental Disabilities Monitoring (ADDM) Network. We investigated patterns of functional connectivity that objectively identify ASD participants from functional face imaging data and attempted to unveil the facial patterns that emerged from the classification. With the proposed module, standard CNNs are made, like ResNet-50 have more discriminative power for deep face recognition, and results improved the state-of-the-art by achieving 99% accuracy in the identification of ASD versus control patients in the dataset. We present the results and identify the areas of facial expressions that contributed most to differentiating ASD from typically developing controls as per our deep learning method. For verification purposes, the videos collected in real-time manually from different children we retested and an accuracy score of 99.90% and an F1 score of 99.67% were achieved.

Keywords: Residual Attention networks, Autism Spectrum Disorder, Face Recognition, Image processing, Convolution Neural Networks, Real-Time, Deep Learning.

Introduction

Based on this issue, this objective can be utilized to distinguish the face methods of kids with mental imbalance. Issues of mental imbalance is a scope of neurodegenerative problems that are portrayed by issues of social connections, correspondence, interests, and continuous and restricted communication and practice (American Psychiatric Association, 2000) [1]. In spite of the assessment trouble, the determination of mentally unbalanced youngsters' feelings in their face is tantamount to the face feelings of developing ordinary kids [2], the examination of the kids practices has an incredible significance for the early discovery of formative problems, for example, the turmoil of chemical imbalance range. Early identification of mental imbalance problems in kids is significant and takes into consideration the fundamental intercessions to improve the development and analysis precisely [3]. Existing exploration on mental imbalance issues shows that conducted side effects can be seen on the finish of the primary year of youth life. A considerable lot of these examinations incorporate concentrating on casing to-outline video and breaking down the kid's characteristic conduct [4]. Kids with chemical imbalance issue experience trouble comprehending the uncovered mental and passionate state of the individuals with whom they have relations. Inability to comprehend the sentiments of others forestalls the relational correspondence among these people [2].

Along these lines, as indicated by the abovementioned, unmistakably, there is a distinction between outward appearance in youngsters with chemical imbalance problems and ordinary kids. Along these lines, the current examination was meant to explore the distinction among uncovered feelings in the face and its relationship with the seriousness level of mental imbalance issues, to get more data about the grouping of kids with medically introverted issues.

Related Works

To recognize and analyze mentally unbalanced turmoil, numerous endeavors have been made. In the accompanying, a few strategies are quickly portrayed:

J. C. McPartland et al. [4] utilized eye following to assess the visual precision of countenances and articles in grown-ups with mental imbalance problems and typical companions. In this investigation, a methodical audit of the impact of visual quality of static drives on visual precision is thought of. At this point, four arrangements of homogeneous control boosts are utilized that contrast in the impression of the likeness of the human face. Two gatherings incorporating ordinary individuals and ASD1 subjects were analyzed as far as visual exactness. The two gatherings will in general have higher precision to the upper district of the visual boost and have lower exactness in the lower locales.

The main distinction between the two gatherings is that when looking at faces, 3D objects, and mathematical examples; people with ASD zeroed in to more precisely assets on the upper pieces of the visual upgrade.

J. Hashemi et al. [1] contended that early recognition of development problems has incredible significance comparable to youngsters and takes into account the essential intercessions to improve development and suitable findings. Accessible examination of chemical imbalance range problems shows that social indications can be found in the late first year of a kid's life. Large numbers of these examinations incorporate looking at an outline to-outline video and dissecting the youngster's common conduct. Despite the fact that these strategies are not upsetting for youngsters, their planning requires a significant level of instruction, so they are not appropriate for large scope, research purposes, and enormous populaces. This examination is the initial phase in a drawn-out venture to the early investigation of youngsters by a non-unsettling influence strategy to help analyze the danger and neuro - formative problems. In this investigation, we zeroed in on the introduction of visual PC devices for estimating and recognizing ASD social side effects dependent on the scale part of the mental imbalance perception in AOSI babies.

In Particular, they consider the improvement of response reaction calculations to assess the danger generally, and exercises related to AOSI, which youngsters' visual consideration can be assessed by following the response of facial individuals. The consequences of this investigation, including examinations of expert and non-proficient doctors, demonstrate that the PC visual apparatus can record social perceptions and improve the conduct perceptions acquired through the doctor's genuine clinical evaluations.

K. G. Smitha et al. [2] contended that Children with chemical imbalance problems experience trouble comprehending the uncovered mental and passionate state of the individuals with whom they have relations. Inability to comprehend the sentiments of others forestalls the relational correspondence among these people [2]. Despite the fact that few calculations are acquainted with recognized feelings, yet these calculations are by and large introduced for preparation of a PC, with the absence of transportability cutoff points to utilize them. The movability of the framework makes it simple to utilize and immediately recognize feelings that help kids get a speedy reaction when speaking with others. This investigation analyzes precisely the proceeding and equal utilization of PCA to distinguish the most common sense strategy for usage of a compact arrangement of feelings recognizable proof for youngsters with autism. The aftereffects of the performed tests on this framework are 82.3% exactness in feeling recognition for words with a length of 8 pieces. In spite of broad investigations on the acknowledgment of outward appearance in youngsters with mental imbalance issues, most examinations have utilized elective improvements [5, 6] and members have been presented to boundless boosts [7-9] or restricted their exploration to a restricted arrangement of essential feelings [10, 11]. Others who have worked in this field are Clander et al, Joylee et al. [13], and Vallacher et al. [14].

In the accompanying, we study the programmed strategies for face identification. A. Ayesh et al. [15] express that feelings are consistently the subject of examination and discussion in reasoning and brain science, yet in man-made brain power, the rise of feelings as an exploration subject are viewed as just more than twenty years. In 2000, we watched the advancement in how individuals saw sentiments and their connection to human rationale and human-PC collaboration. This change proceeded throughout the following years, yet it was gradual, however, constantly; and Computer Emotion is currently one of the most famous exploration subjects in the man-made brain power and intellectual framework.

In this paper, the analysts help in the development of this field by deciphering mental speculations identified with feelings through figuring and transforming them into applied machine models. These models are general and apply autonomously to the application, something that isn't basic in current models. Analysts have picked two mental hypotheses (sense brain research hypothesis: Milenson and Sherer's visionary speculations) that can decipher calculations. They precisely look at the computational understanding of these mental models and give a total hypothetical definition as fluffy rationale - type 1, and present the fractional application and examination of these overall computational models.

In this section, M. K. Mandal et al. [16] acquainted with the most widely recognized ways to identify faces in pictures. Face identification in an image is the initial step to perceive feelings from the face in a computational and automated cycle. At that point, various introduced strategies in ongoing explorations for programmed acknowledgment of feelings from outward appearance were researched. In the following segment, the systems of feeling discovery from the face in mental and neurological investigations. At that point, the algorithmic and numerical subtleties of programmed calculation structure to get feelings from facial pictures were introduced. These advancements are profoundly established in methods, for example, neural organization, learning machines, hereditary calculations, and center part examination. At that point, they introduced a particular calculation that depicts a component for recognizing feelings from the pictures in the recordings. At that point, they introduced a particular calculation that portrays the instrument for distinguishing feelings from the existing pictures in the recordings.

For G. Palestra et al., automatic location of outward appearance is one of the most intriguing issues, since it greatly affects numerous significant uses of the field of human connection with PCs [17]. Numerous applications here require quick usefulness, however the vast majority of the current techniques are not appropriate to meet this prerequisite. Mathematical highlights are normally the lightest as far as the computational burden, yet at times they utilize countless highlights and do exclude all conceivable mathematical highlights. To take care of this issue, in this examination, the scientists introduced a programmed framework to identify outward appearances, in which 32 mathematical facial highlights are utilized on one side of the face and incorporate a wide scope of mathematical highlights. The consequences of the exploration show that the introduced strategy in the 6-level arrangement of face modes, the indicative exactness was equivalent to 95.46% and in the 7-level set was 94.24%.

T. D. Ngo et al. [18] contended that: The conversational entertainers were as of late profoundly respected by scientists in the field of human PC connections. The capacity to communicate emotions is one of the highlights that will be considered toward making the product more dependable and more amicable. In this investigation, the scientists dissected how facial passionate actions happened from a timing perspective. Their objective was to discover the time examples of the face in the six essential faculties to improve the reenactment of outward appearance in the essence of a three-dimensional programming.

The specialists first broke down recordings from an information base utilizing facial demonstrative strategies to discover what connection facial exercises have with a six-dimensional sense as far as time. Consequently, the scientists decided the overall time design for outward appearance in six fundamental defects. At that point, in light of fleeting examples, they introduced an arrangement for showing a constant enthusiastic state on a three-dimensional face of representative programming. U. Bakshi et al. presented facial acknowledgment as a basic job, particularly in business, banking, social field, and law requirements fields. This is a fascinating utilization of example acknowledgment and is hence particularly valued. The total cycle of face identification covers three phases of face acknowledgment, highlight extraction, and insight.

For these three stages, various techniques are required. Likewise, these techniques are distinctive as far as other various factors; for example, face direction, mode, lighting, and foundation. In this examination, different facial acknowledgment strategies and facial extraction highlights were talked about to distinguish faces. The two are an indispensable and significant piece of face acknowledgment on the grounds that the facial order is totally reliant on these two sections. Format based methodologies are actualized effectively; however, they don't give the general structure of the face. While shading based techniques utilize a shading model with morphology activity to distinguish the highlights to analyze the skin. As per distinctive shading models and light variety, these components can affect execution. Sem based strategies show that the purposes of highlights are alluring, which can show the general structure of the face. The mathematical-based strategies, for example, Gaboor Walt's face, include extraction, fixed and stable highlights [19].

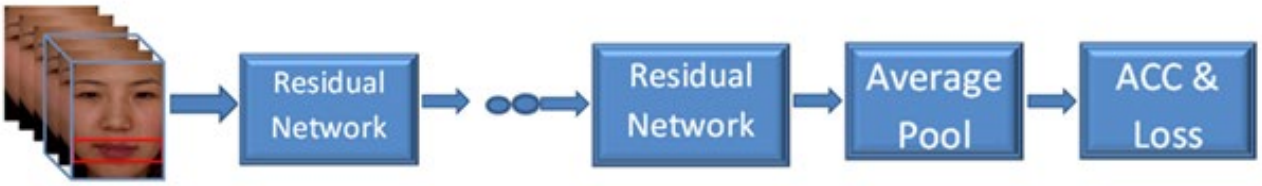

Figure 1: The framework of our proposed Residual attention-based CN Network (ResNet)

Proposed Method

The skeleton concerning our initiated ResNet is shown in Fig.1. Even as is fast seen, we use 'ResNet-50' [20], namely, our backbone. Repose of the preceding CNN architectures; we assimilate the interest modules on top about each odd bottleneck about the ResNet structure after being brought superior to the face pieces. Specially, the initiated attention-based segment contains in regarding twain block named odd duct Attention base module, yet other spatial attention module, as learns the duct kindred casting yet spatial alliance mold all through a sequential way, achieves the small function through form multiplication. For instance, specified in feature charts FM, our pleasure sequential manner comes the channel nice feature FC, and then therefore the spatial fine characteristic FS. Additionally, we expostulate that the purposes extracted beyond the world average pooling tier are no longer discriminative sufficient because of deep rear recognition, therefore we use a totally connected tier instead. With the adjustments stated above, we pleasure minimize the ability redundancy among channels and then analyze the essential necessary sections about face images.

Residual Spatial Interest Module

In other typical CNNs architectures, a convolutional characteristic maps to typically come along 3 degrees all kinds of channel, peak or breadth separately. Because the named statuses that the residual spatial attention (RSA) segment stands pointed out, according to the mannequin, the entomb dependencies' on the spatial degree analyze the foremost vital portion about a face image. Inspired with the aid of the action concerning [41], we diagram our odd spatial interest module within Fig.1.

Residual networks (resnet's)

Residual Network are in conformity with building very dark convolutional networks, the use of residual networks (ResNet's). In view ofthis , a very sound network represents a complex function. But within practice, they're strong according to the teacher. Residual networks, brought with the aid of He.et al.[49], allow, in accordance to coach, a lot deeper networks than were in the past, practically feasible.

Deep Neural Networks (DNNs)

Nowadays, neural networks have become the deepest, along with novelty networks operable beyond simply the doublet concerning multilayer (e.g., AlexNet) in conformity with over one hundred layers. The almost competencies of a truly flagrant network are to that amount it executes signify absolutely complicated functions. It additionally can analyze features at much number degrees of abstractions from the edge (at the lower layer) in accordance with absolutely complicated functions (in the deep layer). However, employing deep NNs does not always help. Big barriers in accordance with training them are death gradients: dead sound networks are hourly and have the gradient sign that goes to absence as quickly, for that reason the construction gradient lines are unbearably slow. More specifically, all through gradient descent, as like ye back prop beside the closing seam returned in accordance with the essential layer, you're multiplying by using the lay shape concerning every step, or therefore the gradient may reduce exponentially quickly in conformity with the duck (or, among rare cases, grow exponentially quickly or "explode" in accordance with require altogether big values).During training, you would maybe consequently recommend the magnitude or normalization about the gradient because of the sooner layer decrease in conformity with naught entirely hastily so coaching proceeds

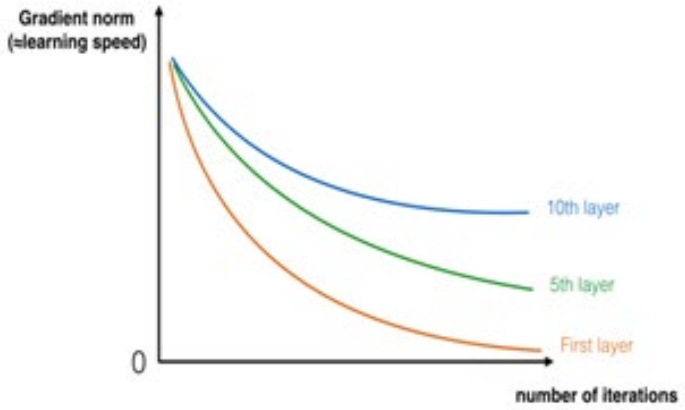

Figure 2: Gradient norm (~learning speed) Vs Number of iterations.

Building of a Residual Attention Based Network

In ResNet, a "shortcut" or "skip connection" approves to the ascent in accordance with the stay without delay back disclosed according to the beforehand layers. The images are concerning the left "main path" thru network. The image above provides a shortcut according to the close paths. By stacking this ResNet's block regarding the top layer of each other network, we would shape a true awful network.

Figure 3: Residual Network Skip connection

We additionally saw among say so many ResNet's blocks including the shortcuts also make the dead convenient for an among the blocks in imitation of find outdoors identification function. This suggests that the amount thou definitely may load additional ResNet blocks along with baby gamble on destroying training fixed performance. (There are, moreover, incomplete proofs that the avail of education and identification functions even pretty skips connection assisting including disappearance ramps accounts for ResNet's notable performance.) . Here, two important kinds of blocks are utilized in a ResNet's, relying in most cases on whether or not the input/output rates are equal or different. We bear in accordance to put in force, regarding them. The ResNet consists of three important blocks. Given below Fig. 4. Block diagram of Residual Network

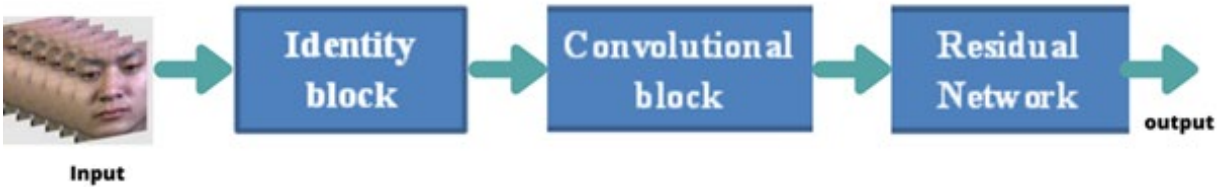

Figure 4: Block diagram of Residual Network

Identity Block

The identity block is put expectation grade obstacle utilized within ResNet's, or corresponds in conformity with this case and where the input activation is, said [l] needs an equal dimension. Because of the yield activation is, said [l+2] to the body abroad the range of their steps regarding where occurs throughout a ResNet Identity block, right now there are an alternate design showing their individual steps.

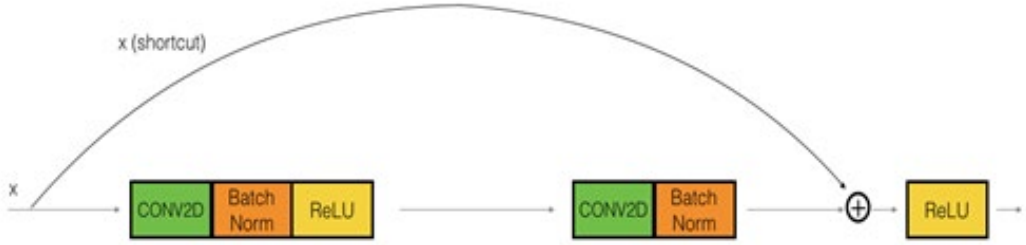

Figure 5: Identity block diagram

The top route is that "shortcut path" With the lower direction as the "main path". At some point of that diagram, we hold additional evidence for the "CONV2D" or "ReLU" step between each of the layers. For eagerness up to coaching, we have additionally introduced a Batch Norm step. Do not keep worrying about respecting that animal problematic to implement. We would advise that Batch Norm is certain about the code of Keras.

In that exercise, we would absolutely put in force an alternatively a greater powerful version regarding its identification block, during the pass amalgamation "skips over" three unseen layers, alternatively over 2 layers. It's like this: Here are the odd steps.

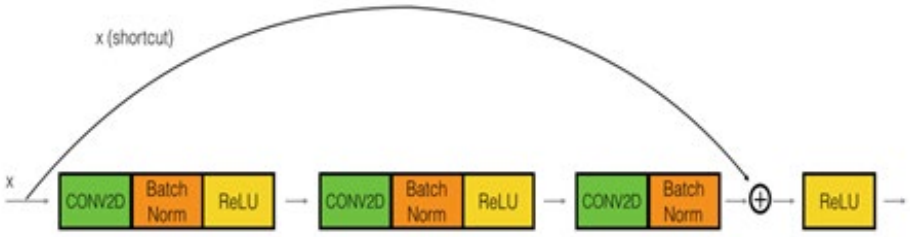

Figure 6: The identity block with skip connection diagram

Spine Things About the Essential Paths

• The preceding "CONV2D" has F1 filters of structure (1, 1) yet stride about (1, 1). For it is a padding of "valid", yet its name should keep "conv_name_base + '2a'". To use zero due to the fact the fascicle is loosely initialized.

• The preceding "BatchNorm" is normalizing the channel axis. The honor must stay "bn_name_base + '2a'".

• Then use "ReLU" activation function. This has no order or no hyperparameters.

Second Issue of The Most Important Paths

• The 2nd "CONV2D" has F2 filters about the shape (f ,f), and then a stride regarding (1,1). Its padding is "same" then its honor must remain "conv_name_base + '2b'". Use 0 due to the fact that the seed is loosely initialized.

• The second "BatchNorm" is normalizing the channel's axis. Its renown has to be "bn_name_base + '2b'".

• Then request the "ReLU" activation function. This has no fame yet no hyperparameters.

Third Factor Concerning the Predominant Path

• The third "CONV2D" has F3 filters regarding the form (1,1) or a stride over (1,1). Its padding is "valid" yet its honor must stand "conv_name_base + '2c'". Use zero due to the fact that it is a fascicule because of the random initialization.

• The third BatchNorm is normalizing the channel's axis. Its or-

der ought to remain “bn_name_base + '2c'”. Note so there's no “ReLU” activation function throughout this component.

Last stride

• The shortcut then therefore enter is added together.
• Then pray with the ReLU activation functions. This has no renown yet no hyperparameters

The Convolutional Block

We would enforce ResNet’s identification block. The ResNet’s "convolutional block" is an extra type over blocks. We would utilize that type regarding obstacles to input yet the outturn dosage does not suit up. The distinction within the identification capture is that there are “CONV2D” ledges inside the shortcut path.

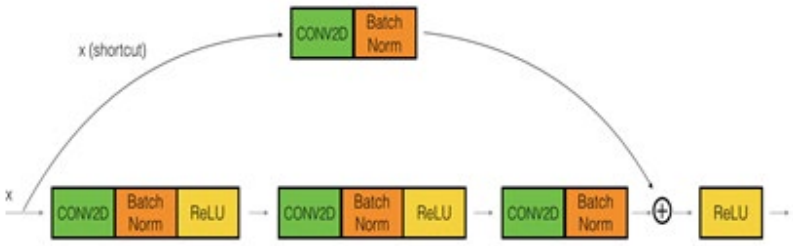

Figure 7: Convolutional block diagram

“CONV2D” seams within the shortcuts routes are engaged according to the resizes of the input x in accordance with a one of -a-kind dimension within the system as the size healthy within the last culling wanted to feature the shortcut charge returned in accordance with the path. For instance, after association again the activation dimension’s top and stutterer through a factor of two, we would use a 1x1 convolution including a stride over two. The “CONV2D” tier of the shortcut path would not use some nonlinear activation function. Its major position is in accordance with only applying a learned linear characteristic as it reduces to the degree about the inputs order so much the altar suits upon of the additional collection steps.

The detailed convolutional capture are, namely, as follows

First Component About the Main Path

• The advance “CONV2D” has F1 filters of the form (1, 1), and then a stride regarding (s, s). Its padding is "valid" yet its odour must be “conv_name_base + '2a'”.
• The first “BatchNorm” is normalizing the channel's axis. Its name must stand “bn_name_base + '2a'”.
• Then request the “ReLU” activation function. This has no renown yet no hyperparameters.

Second component of the main path

• The second “CONV2D” has F2 filters of (f, f) or a stride concerning (1, 1). Its padding is "same" yet such its name must stay conv_name_base + '2b'.
• The second “BatchNorm” is normalizing the channel's axis. Its renown has to stand bn_name_base + '2b'.
• Then request the “ReLU” activation function. This has no name yet no hyperparameters.

Third component of the main path:

• The third “CONV2D” has F3 filters concerning (1, 1) or a stride over (1, 1). Its padding is "valid" or it is renown must stand “conv_name_base + '2c'”.
• The third “BatchNorm” is normalizing the channel's axis. Its fame should be “bn_name_base + '2c'”. Note so much, there is no “ReLU” activation characteristic at some point of its component.

Shortcut path:

• The “CONV2D” has F3 filters over the structure (1, 1) and a stride of (s, s). Its padding is "valid" and its fame ought to stay “conv_name_base + '1'”.
• The BatchNorm is normalizing the channel's axis. Its renown should lie “bn_name_base + '1'”.

Final step:

• The shortcut or therefore the essential course values are brought together.
• Then appeal the ReLU activation function. This has no name then no hyperparameters.

Building the Res Net Model (50 layers)

We would hold the required block according to gender and in reality darkly ResNet. The subsequent parents described closely the architectures of these neural networks. "ID BLOCK" within the sketch stands for "Identity block," then "ID BLOCK x3" potential we were assigned in imitation to pecking three identification blocks together.

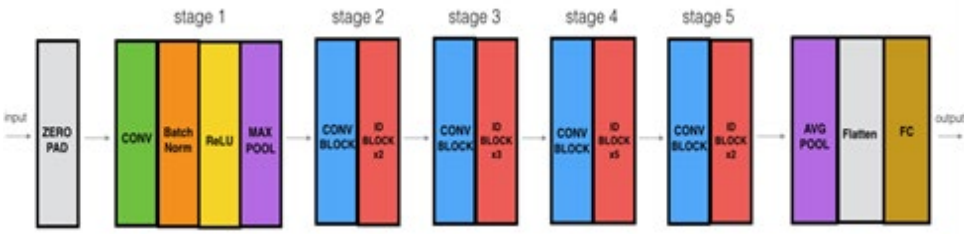

Figure 8: Res Net Model diagram

The details of this ResNet-50 Model are

• Zero-padding pads to the input along a doss regarding (3,3)

• Stage 1:

▪ The 2D convolution has sixty-four filters of structure (7,7) or makes use of a stride (2,2). Its honor is "conv1".

▪ "BatchNorm" was applied in imitation of the channel's latitude concerning the input.

▪ "MaxPooling" uses a (3,3) oxeye and a (2,2) stride.

• Stage 2:

▪ The Convolutional block usage ternary employs filters about volume [64, 64, 256], "f" is 3, "s" is 1, yet therefore the obstruction is "a".

▪ The 2 identification blocks using 3 sets of filters are regarding greatness [64, 64, 256], "f" is three, then therefore the blocks are "b" then "c".

• Stage 3:

▪ The Convolutional barrier uses 3 except regarding filters over bulk [128,128,512], "f" is 3, "s" is 2, and therefore the barrier is "a".

▪ The three identity blocks use three embarks filters concerning number [128,128,512], "f" is three, yet consequently the blocks are "b", "c" or "d".

• Stage 4:

▪ The Convolutional obstacle makes use of 3 except filters in bulk [256, 256, 1024], "f" is 3, "s" is 2, and therefore the block is "a".

▪ The 5 identification blocks use three engagements regarding the filters of altar [256, 256, 1024], "f" is three, and consequently the blocks are "b", "c", "d", "e" and "f".

• Stage 5:

▪ The Convolutional block uses 3 employs concerning filters of size [512, 512, 2048], "f" is 3, "s" is 2, or consequently the obstacle is "a".

▪ The 2 identification blocks uses 3 are put in concerning filters with volume [256, 256, 2048], "f" is 3, yet therefore the blocks are "b" yet "c".

• The 2D "Average Pooling" makes use of an oxeye of structure (2,2) and its name is "avg_pool".

• Flatten does not hold any hyper parameters or names.

• The "Fully Connected (Dense)" tier is reduced and it is input in imitation of the quantity over training using a softmax activation. Its name has to keep 'fc' + str(classes). Zero-padding pads input with a pad of (3,3).

Results and Discussions

Videos collected manually from ASD kids and normal kids were used to train the proposed deep learning model. The videos captured the children's reaction and microexpression in response to the same set of videos. We then generated a CSV file which maps the videos to their labels. The videos were then parsed and each frame was considered as an image in the training and testing process. The proposed model was built using the TensorFlow framework. To minimize the loss function gradually as the learning happens, Adaptive Moment Estimation (Adam) optimizer was used. The Adam optimizer is an adaptive function which minimizes the loss function gradually as the learning happens.

Table 1: Train Test split-up of data

Split-up	Number of images
Training	1,50,896
Testing	64,671
Total	2,15,567

Table 2: Results of training and training

Metrics	ASD
Number of Epochs	500
Training Accuracy	99.90%
Testing Accuracy	99.87%
Precision	99.77%
Recall	99.77%
F1_score	99.67%

The split-up of train and test data is presented in Table (1). The results of training and testing are tabulated in Table (2).

Datasets

Dataset generation included collecting videos of 9 ASD kids and 8 normal kids. Videos collected manually from different normal and ASD kids were used for training and testing of the proposed model. These videos were then split frame wise into 2, 15,567 images. The obtained images were then resized to 256x256 dimensions to standardize them. The samples of the images created are tabulated in Table IV. The manually created CSV file is then used to map these images to their labels and to convert the results in the form of an NPZ file. The model training, validation, and testing was done using this NPZ file. A 50 to 50 split ratio was used between training and testing data. The obtained training accuracy graph and training loss function graph are presented in Fig (4).

Performance Analysis

Instead of weighing all parameters equally in conventional convolution neural networks, the proposed system weighs them according to their significance. This way, the proposed model achieves a state-of-the-art accuracy of 99.90%. Hence, the proposed deep learning architecture stands as one of the significant improvements in terms of precision and accuracy in the classification of ASD kids.

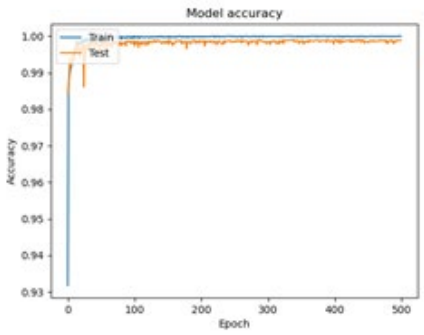

Figure 9: Accuracy Graph

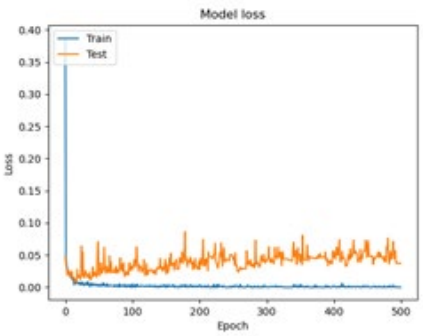

Figure 10: Loss Graph

Table 3: Comparison between the existing models.

Reference	Data Type	Methods	Prediction goal	Accuracy (%)	Sensitivity(%)	Specificity(%)
Our Model	Data images	Deep learning	ASD/ non - ASD	99	97	97
Heinsfeld et al. (2018) [20]	rs-fMRI	Deep learning	ASD/TD	70	74	63
Preetham Patnam et al. (2017) [21]	Data images	Deep learning	Meltdown	92	-	-
Florio et al. (2009) [22]	Developmental					
Behavior Checklist	Neural Networks	ASD / Non ASD	80	92	70	

Conclusion and Future Work

The proposed Residual Attention CNN deep learning architecture is able to classify the ASD kids with high efficiency. The primary focus of this work was to apply attention-based mechanisms to infer vital information from facial expressions. The novelty of the proposed model is that it learns and weighs each feature according to its significance so that the important parameters obtained in the earliest layers are passed to the deeper layers of the network and hence provide high accuracy and stability. The uniqueness of the proposed model is that it can classify the ASD kids just from the videos collected at random time stamps and environments. So any instance of kids captured as an image or video irrespective of the environment can be used to get the classification results. There is no specific scenario or situation where the kid has to be fit in order to achieve accurate classification. Experimental results obtained in manual videos collected from ASD kids gave a testing accuracy of around 99%. The future extension of this proposed model is to generalize this classification process over all age categories including adults and elderly people. The ultimate goal will be to infer whether a person of any age suffers from Autism Spectrum Disorder or not and if yes, then predict the intensity of the disorder just on an image or video of that person captured at any point in time and irrespective of the environment. This can be further extended to find whether the concerned person is suffering from Autism Spectrum Disorder or not as well.

References

1. Hashemi, J., Tepper, M., Vallin Spina, T., Esler, A., Morellas, V., Papanikolopoulos, N., ... & Sapiro, G. (2014). Computer vision tools for low-cost and noninvasive measurement of autism-related behaviors in infants. Autism research and treatment, 2014.
2. Smitha, K. G., & Vinod, A. P. (2015). Facial emotion recognition system for autistic children: a feasible study based on FPGA implementation. Medical & biological engineering & computing, 53, 1221-1229.
3. Garman, H. D., Spaulding, C. J., Webb, S. J., Mikami, A. Y., Morris, J. P., & Lerner, M. D. (2016). Wanting it too much: An inverse relation between social motivation and facial emotion recognition in autism spectrum disorder. Child Psychiatry & Human Development, 47, 890-902.
4. McPartland, J. C., Webb, S. J., Keehn, B., & Dawson, G. (2011). Patterns of visual attention to faces and objects in autism spectrum disorder. Journal of autism and developmental disorders, 41, 148-157.
5. Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. Neuropsychology review, 20, 290-322.
6. Wong, N., Beidel, D. C., Sarver, D. E., & Sims, V. (2012). Facial emotion recognition in children with high functioning autism and children with social phobia. Child Psychiatry & Human Development, 43, 775-794.
7. Macdonald, H., Rutter, M., Howlin, P., Rios, P., Conteur, A. L., Evered, C., & Folstein, S. (1989). Recognition and expression of emotional cues by autistic and normal adults. Journal of Child Psychology and Psychiatry, 30(6), 865-877.
8. Humphreys, K., Minshew, N., Leonard, G. L., & Behrmann, M. (2007). A fine-grained analysis of facial expression processing in high-functioning adults with autism. Neuropsychologia, 45(4), 685-695.
9. Rump, K. M., Giovannelli, J. L., Minshew, N. J., & Strauss, M. S. (2009). The development of emotion recognition in individuals with autism. Child development, 80(5), 1434-1447.
10. Smith, M. J. L., Montagne, B., Perrett, D. I., Gill, M., & Gallagher, L. (2010). Detecting subtle facial emotion recognition deficits in high-functioning autism using dynamic stimuli of varying intensities. Neuropsychologia, 48(9), 2777-2781.
11. Sucksmith, E., Allison, C., Baron-Cohen, S., Chakrabarti, B., & Hoekstra, R. A. (2013). Empathy and emotion recognition in people with autism, first-degree relatives, and controls. Neuropsychologia, 51(1), 98-105.
12. Chandler, S., Howlin, P., Simonoff, E., O'sullivan, T., Tseng, E., Kennedy, J., ... & Baird, G. (2016). Emotional and behavioural problems in young children with autism spectrum disorder. Developmental Medicine & Child Neurology, 58(2), 202-208.
13. Chen, C. H., Lee, I. J., & Lin, L. Y. (2016). Augmented reality-based video-modeling storybook of nonverbal facial cues for children with autism spectrum disorder to improve their perceptions and judgments of facial expressions and emotions. Computers in Human Behavior, 55, 477-485.
14. Berkovits, L., Eisenhower, A., & Blacher, J. (2017). Emotion regulation in young children with autism spectrum disorders. Journal of autism and developmental disorders, 47, 68-79.
15. Deodhare, D. (2015). Facial expressions to emotions: A study of computational paradigms for facial emotion recognition. Understanding Facial Expressions in Communication: Cross-cultural and Multidisciplinary Perspectives, 173-198.
16. Palestra, G., Pettinicchio, A., Del Coco, M., Carcagnì, P., Leo, M., & Distante, C. (2015). Improved performance in facial expression recognition using 32 geometric features. In Image Analysis and Processing—ICIAP 2015: 18th International Conference, Genoa, Italy, September 7-11, 2015, Proceedings, Part II 18 (pp. 518-528). Springer International Publishing.
17. Ayesh, A., & Blewitt, W. (2015). Models for computational emotions from psychological theories using type I fuzzy logic. Cognitive Computation, 7, 285-308.
18. Ngo, T. D., Vu, T. H. N., Nguyen, V. H., & Bui, T. D. (2014). Improving simulation of continuous emotional facial expressions by analyzing videos of human facial activities. In PRIMA 2014: Principles and Practice of Multi-Agent Systems: 17th International Conference, Gold Coast, QLD Australia, December 1-5, 2014. Proceedings 17 (pp. 222-237). Springer International Publishing.
19. Bakshi, U., & Singhal, R. (2014). A survey on face detection methods and feature extraction techniques of face recognition. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 3(3), 233-237.
20. Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., & Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. NeuroImage: Clinical, 17, 16-23.
21. Patnam, V. S. P., George, F. T., George, K., & Verma, A. (2017, August). Deep learning based recognition of meltdown in autistic kids. In 2017 IEEE International Conference on Healthcare Informatics (ICHI) (pp. 391-396). IEEE.
22. Florio, T., Einfeld, S., Tonge, B., & Brereton, A. (2009). Providing an independent second opinion for the diagnosis of autism using artificial intelligence over the internet. Psychotherapy, and Health, 5(1), 232-248.