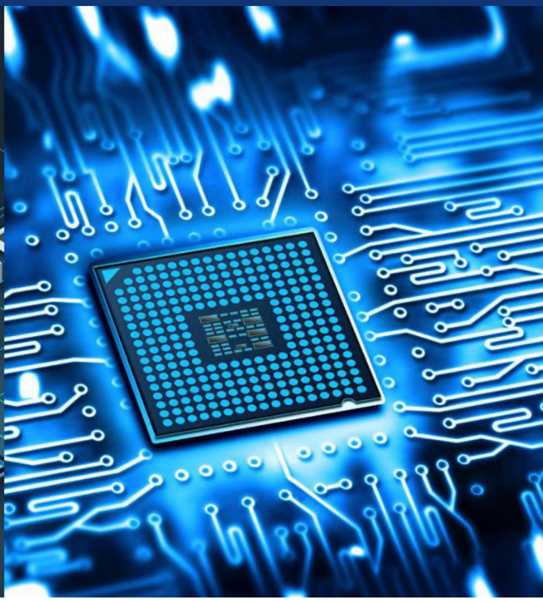
IJIRCCE

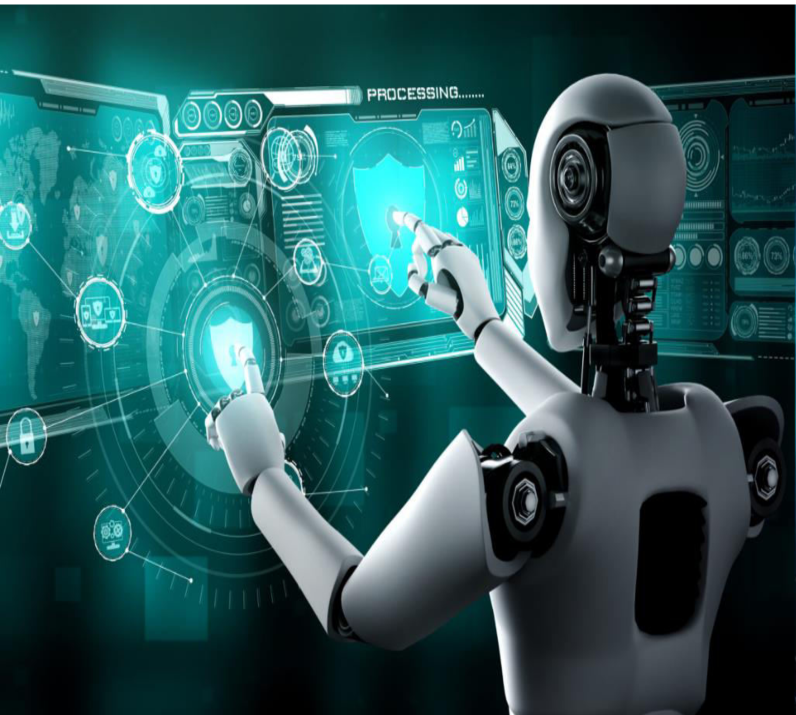
ISSN(O): 2320-9801

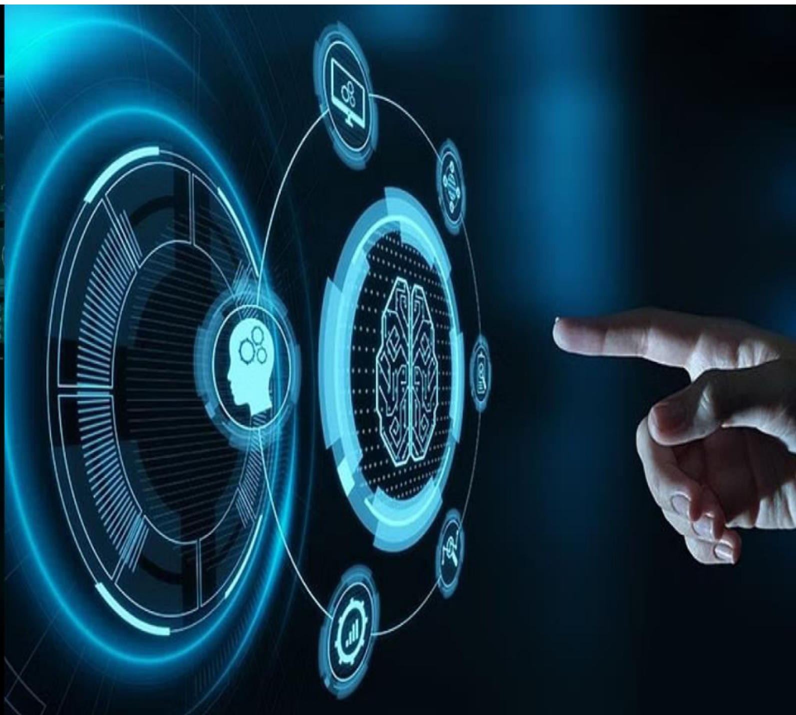
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 9.927

Volume 14, Issue 8, August 2026

www.ijircce.com | ijircce@gmail.com | +91-9940572462 | +91 63819 07438

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408007

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Hybrid Multimodal Autism Spectrum Disorder Screening Framework

Swathi Iddum, Aki Sai Surya Kamal Tarun, Paidi Lahari, Nallana Kavya Sri, Vemula Sai Pavan

Department of Computer Science & Engineering – Data Science, Anil Neerukonda Institute of Technology and Scienes, Visakhapatnam, India

ABSTRACT: Autism Spectrum Disorder must be detected at an early stage to provide help without any delay. A multimodal approach combining facial image and behavioral data is proposed in this study for improved early detection of ASD. To obtain more effective results, this approach uses a combination of methods like EfficientNetB3, MultiLayer Peceptron (MLP), Hybrid Feature Fusion and XGBoost. This framework combines behavioral data with facial image data to retrieve informative patterns related to ASD and enhance classification accuracy. This approach has achieved an accuracy of 88.31 percent after training, which is higher than the other approaches which existed previously. A clear pattern is observed that combining multiple methods improves accuracy rather than using a single approach. It also indicates that the combination of diverse models improves the performance of early-stage ASD detection.

KEYWORDS: Autism Spectrum Disorder (ASD), Deep Learning, Convolutional Neural Network (CNN), EfficientNet, Multimodal Learning, Feature Fusion, XGBoost, Facial Image Analysis

I. INTRODUCTION

People sometimes refer to it as Autism Spectrum Disorder or ASD for short. Doctors think of autism as a kind of condition. This way of thinking is present from birth. It is not affected by how someone is raised or by any significant life events. Brains that are wired this way do not change just because someone was raised in a way. Each person with autism is different. Some people with autism might start speaking while others might never speak at all. Sometimes people with autism are really happy. A few people with autism can manage on their own without needing help from others. Others might find that they need help with tasks. This range of needs is why autism is referred to as a spectrum. Earlier research tried to use photos to diagnose autism using something called VGG16 and VGG19[1]. These methods were not very good at catching small visual clues. They would often just repeat what they had learned before.

If you look closely at how someone with autism acts, you can see clues. Childhood is a time when these signs often appear although some signs can be stronger than others. People with autism might do things like spin, hit their head or wave their arms. They might have trouble talking or making eye contact with others. Recently there have been reports that more children are being diagnosed with autism. In 2012 about 15 out of every 1000 children had autism. Now that number is 33 out of every 1000 children[2]. Symptoms of autism usually start to appear when a child is two years old. This increase in autism diagnosis has gotten a lot of attention from researchers. They think things like the environment and the mother's health might be causing more children to have autism [3].

If autism is caught early, it can be easier to deal with. Care can be more effective when signs are noticed sooner. Each person with autism is different and will react in their own way. Autism is not about how someone behaves. It is about how their mind processes what they see and hear [4]. Faces can give clues about autism because the expressions on someone's face can show how they communicate without using words. People with autism can behave in different ways. It affects how people think and process information. Now computers are getting better at helping us with things like medicine. Because of this using intelligence to spot health issues is becoming more popular in scientific research. This is especially true when it comes to looking at evidence for diagnosis. Of using the old VGG16 and VGG19 methods this research uses something called EfficientNet to look at pictures of children's faces and try to determine if they have autism. This method is better because it can see the expressions on a child's face when they are looking forward. The researchers made some changes to the model, which helped it to learn without memorizing things. This made the system better at finding clues in patterns that are tied to autism. Earlier methods were only able to diagnose autism about 82.55% of the time[1]. This new method is able to diagnose autism about 88.31% of the time. What is notable about this method is how well it can find patterns in faces to diagnose autism. It is better than the methods and can produce more consistent results.

II. LITERATURE SURVEY

Nowadays, work on spotting Autism Spectrum Disorder is not limited to one method. It brings together face data, behavior, and body signals, unlike older CNNs that mostly analyze image pixels [1]. They can catch a few signs of autism but miss the more detailed ones. They also struggle because ASD can look very different from person to person. Combining different types of data usually gives better results. This project looks at computer-detected face features, main points on the face, and behavior survey responses. Recent findings indicate that using a mix of related attributes beats looking at single attribute on their own when spotting signals [5].

EfficientV2S and EfficientB3 are used together to get detailed visual features from facial images. They perform better than older models like VGG16 [1] and MobileNet [5], but don't require extra computing power. These models are good at finding detailed patterns in faces, including autism-related signs, even when the dataset is small. Using MediaPipe FaceMesh, the system tracks 478 points on each face in 3D. Studies show how children on the spectrum sometimes have unique facial features, like wider eyes or a shorter middle face [3]. By converting images into coordinate maps, the system can see slight changes in the face shapes, improving results and understanding.

The system takes answers from people and adds behavior clues from face scans. These answers are turned into a 1-10 score and is processed by Extreme Gradient Boosting (XGBoost) [2]. It works effectively with structured data and does better than older models like SVM and RandomForest [7] in spotting small patterns. It works well even with uneven groups. Looking at faces and behavior at the same time makes the system more accurate, reducing both false positives and missed cases.

CNN outputs and facial landmarks are given as input to the MultiLayer Perceptron (MLP). Noisy data is handled by using Batch Normalization [5] along with Dropout which is used to improve stability. The system uses a sigmoid gate at the end to turn outputs into probabilities, making decisions clearer. It also helps in enhancing the older models. The system uses EfficientNet for visual patterns, MediaPipe for facial points, and XGBoost for behavior. Together, they form a more powerful, clear, and precise model that doesn't depend on skin or other demographic features [8].

III. METHODOLOGY

The process begins with data preparation and cleaning. Then the system learns from it. Once training is done, validation is used to check its performance, and testing shows how it works on unseen data.

3.1 Dataset

The study uses a dataset named autism-image-data [1], which you can access for free on Kaggle. It has training, testing, and validation groups, with images of both autistic and non-autistic individuals. Every image file has a name starting with Autistic or Non-Autistic, then a number, and ends with .jpg. The dataset contains 2940 images. The training data includes 1270 autistic and 1270 non-autistic images, while the validation set has 50 images from each group. There are 301 images in the test set. The dataset includes people from many parts of the world, and all races are fairly represented. Figure 1 displays samples from both groups.

Table 1. Dataset Distribution

Dataset Partitioning	Autistic	Non-Autistic	Total
Training	1270	1270	2540
Validation	50	50	100
Testing	150	151	301
Total	1470	1471	2941

Autistic 34.jpg Non-Autistic 43.jpg

Fig. 1: Illustration of image format

3.2 Data Preprocessing

Here, the system processes the data in two parallel steps. One step looks at spatial features, while the other studies facial landmarks. Every image is marked as 1 for autistic or 0 for non-autistic

1)Image Augmentation and Normalization

The facial image is changed a lot, so the system learns the basic patterns instead of remembering specific details of the image. The system divides each color value by 255, which results in the pixel value to be in the range 0 and 1 The facial image is also slightly turned to the left or right up to 25 degrees to show how the image would look when someone moves their head. Some facial images are also moved a little to the side or up and down by 10 percent. The facial image is not always in the same position. The system also zooms in or out of the face picture by 20 percent, which helps the system to recognize face pictures that are taken from different distances, from the camera. Finally, each image is made to fit a size of 224 × 224 pixels, irrespective of its original shape.

2)Geometric Landmark Extraction

The system uses MediaPipe FaceMesh to look at the shape of things at the time. First it changes the pictures, from BGR to RGB and then it finds 478 points on the face. Each point has three numbers that say where it is. After that it puts all these numbers together in a line. This helps the Multilayer Perceptron classifier understand how different parts of the face work together. A scaling step is then applied to normalize the data by centering the values around zero while maintaining a uniform distribution across features[5]. The Multilayer Perceptron classifier can look at the eyes, nose and lips, see how they are related to each other. The MediaPipe FaceMesh and Multilayer Perceptron classifier work together to make this happen.

3) Questionnaire Data Cleaning

From the autism survey, responses labelled A1 through A10 are extracted first. In the dataset, Yes and No answers are changed into numerical form, with Yes assigned the value 1 and No assigned to the value 0[2]. These behavior markers are linked with face related traits. The autism survey responses and face traits are then combined. This combined autism survey data is used for the classification process of autism.

3.3 Training the Model

1)Training the Image-based Model

ASD detection in this system is powered by both deep learning and machine learning. The model is trained for around 30–40 cycles, using batches of 16–32 samples. Adam optimizer is used by the model with a learning rate of 1/10000. If progress becomes slow, the learning rate is decreased by 70% to improve results. Binary cross-entropy is chosen since it suits yes/no decisions. The training setup uses EfficientNetV2S and EfficientNetB3 to extract deep features from the data. The system begins by using EfficientNet to capture key details from images. It improves efficiency by scaling depth, width, and image size together rather than only increasing layers. It also uses depthwise separable convolutions, which make the model lighter without losing useful information. The first 40 layers are not updated during training to preserve basic visual understanding [3]. The final layers adapt to capture ASD-related facial features. Global Average Pooling then reduces the output size, helping the system ignore small shifts in position or pose.

The model uses dense layers with ReLU to learn patterns. Normalization helps maintain consistency, and dropouts randomly skip some parts during training to prevent the model from overlearning. The system combines visual features from the CNN with motion data and behavior survey answers. A final output unit uses a sigmoid function to give a result near 0 or 1 for classification.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

2)Training Questionnaire Models with XGBoost
The system uses survey answers A1 to A10 in an XGBoost model with 300 trees, each up to 4 layers deep [2]. The dataset is divided to maintain equal representation of both classes. This design helps the model understand complex patterns without capturing unnecessary noise.

3)Combining Choices and Improving Results
The final decision is made by combining image-based analysis and behavior inputs. The CNN gives a score using sigmoid, and XGBoost makes its prediction using log-loss as the learning measure. The system combines image analysis with behavior data. XGBoost gives a probability score using sigmoid, while CNN makes its own prediction based on training. Both results are combined, so facial patterns are supported or refined by survey data. Accuracy is thus improved and the errors are reduced.

3.4 Testing the Model
The working of the designed model is checked during validation. For this a group of 301 images are used in which half of them show autistic children and half do not along with the behavior ratings. First, both the models which were built using MultiLayer Perceptron and XGBoost analyze individually and later the results of both the models are combined to give the final result. Balanced score appear after analyzing correct guesses, false alarms and missed cases. The classification depends heavily on the images. To prevent overfitting, validation loss is used to control training. This helps the model produce stable results like real autism screenings.

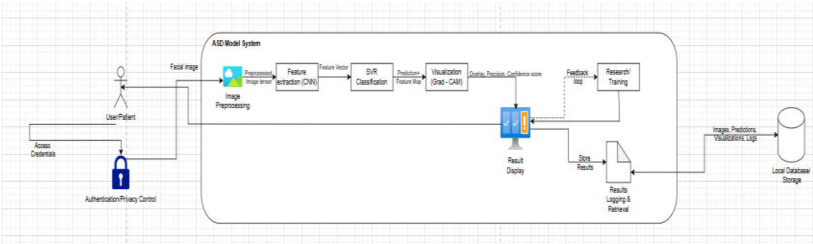


Fig. 2. Architecture of the System

IV. RESULTS AND ANALYSIS

4.1 Model Performance Evaluation
The training and evaluation of the proposed ASD detection system is done using facial image data along with behavioral questionnaire inputs. The system achieved a test accuracy of 88.31% after 40 training epochs, indicating strong classification ability in distinguishing between ASD and Non-ASD children.

A test loss of 29% was observed which shows that the model performed well even with the unseen data and generalized the data effectively. Model accuracy and loss curves give us the learning behavior of the model across epochs.

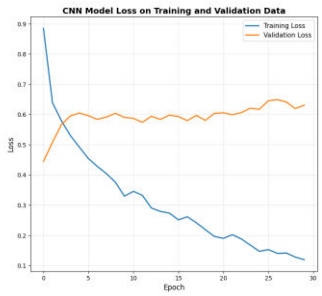
The change in model accuracy during training and validation is shown in Fig. 3. The accuracy increases steadily. As epochs increase, validation accuracy settles, indicating reduced overfitting and steady learning.

The trends of how much our model loses when its training and validation are shown in Fig. 4. Consistent model optimization is reflected by falling training loss and stable validation loss.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



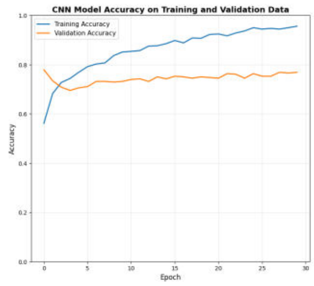


Fig. 3. Model Accuracy Curve **Fig. 4.** Model Accuracy Curve

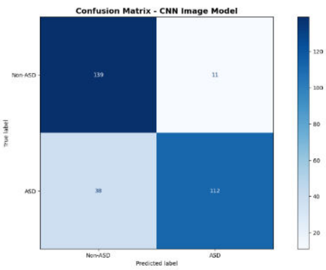
4.2 Confusion Matrix Analysis

Fig. 5 and Fig. 6 shows us the classification performance for both the CNN image model and the questionnaire-based model.

For the CNN image model, the confusion matrix reveals that:

The confusion matrix of the image-based model shows that most of the Non-ASD samples were classified correctly. There were only a few misclassifications. Balanced performance is maintained by the model across both the classes reducing false positives and false negatives.

The questionnaire-based model explains amazing classification behavior, with very few wrong predictions. The effectiveness of behavioral features in ASD detection is confirmed by the high number of true positives and true negatives. The confusion matrices highlight the model's consistency and dependability with real-world data.



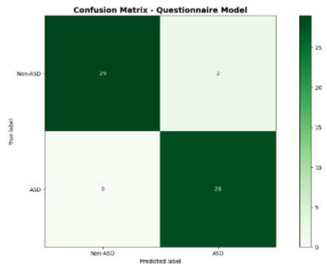


Fig. 5. CNN Model **Fig. 6.** Questionnaire Model

V. CONCLUSION AND FUTURE SCOPE

This work introduces a deep learning framework for identifying Autism Spectrum Disorder (ASD) through facial images and questionnaire-based behavioral data. The previous models either used only facial images or questionnaire-based behavioral data but not together. This system clubs both these approaches into a single multimodal system. The system takes the image as well as responses from the user then they are given as input for their respective models which were built using XGBoost and MultiLayer Perceptron. Each model provides its output and then these outputs are joined together to provide the final output. This multimodal is trained using 40 epochs which results in a test accuracy of 88.31%. This multimodal system offers early screening which is affordable and effective. Future enhancements may include the incorporation of multimodal data sources, such as speech-based inputs, and the use of larger and more diverse datasets to further improve classification performance and robustness. The proposed model outperforms existing CNN based approaches such as VGG16[1] and MobileNet [6].

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

REFERENCES

[1] Chandra, R., Agarwal, S., Tiwari, S., Syafrullah, M., Kumar, A., Adiyarta, K.: Autism spectrum disorder detection using autistic image dataset. In: 2023 10th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), pp. 54–59 (2023)

[2] Shetty, K., Murthy, A., Savitha, G.M.H., Sanjana, Prathwini: Autism prediction using machine learning. In: Proceedings of the 6th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) (2025)

[3] Nathiya, S., Arivazhagan, N.: Early detection of autism spectrum disorder in children using machine learning techniques and DenseNet model. In: Proceedings of the 6th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) (2025)

[4] Haque, N., Islam, T., Erfan, M.: An exploration of machine learning approaches for early autism spectrum disorder detection. Healthcare Analytics 100379 (2025)

[5] Chaitanya, A.R., Muthaiah, D., Sanjai, S.H., Arjun, M., Sb, H., Yugander, P., Jagannath, M.: Autism spectrum disorder detection using attention-based CNN and ML classifiers. Procedia Computer Science 258, 4216–4227 (2025)

[6] Dodia, S., Meshram, V., Kasle, J., Gomase, S., Amrit, H., Sarse, R.: Autism spectrum disorder (ASD) detection from facial images using MobileNet. In: 2024 IEEE 9th International Conference for Convergence in Technology (I2CT) (2024)

[7] Babu, T., Nair, R.R., Sb, K., Mahapatra, S.: Enhancing early diagnosis of autism spectrum disorder using a Bayesian-optimized random forest algorithm. Procedia Computer Science 258, 1576–1585 (2025)

[8] Ashrafi, A.F., Kabir, M.H.: Enhanced graph convolutional network with Chebyshev spectral graph and graph attention for autism spectrum disorder classification. arXiv preprint arXiv:2511.22178 (2025)

[9] Kasri, W., Himeur, Y., Copiaco, A., Mansoor, W., Albanna, A., Eapen, V.: Hybrid vision transformer-Mamba framework for autism diagnosis via eye-tracking analysis. arXiv preprint arXiv:2506.06886 (2025)

[10] Simeoli, R., Rega, A., Cerasuolo, M., Nappo, R., Marocco, D.: Using machine learning for motion analysis to early detect autism spectrum disorder: a systematic review. Review Journal of Autism and Developmental Disorders (2024)

[11] Shrivastava, T., Chaudhari, H., Singh, V.: Leveraging machine learning for early autism detection via INDT-ASD Indian database. arXiv preprint arXiv:2404.02181 (2024)

[12] Liu, X., Hasan, M.R., Gedeon, T., Hossain, M.Z.: MADE-for ASD: a multi-atlas deep ensemble network for diagnosing autism spectrum disorder. arXiv preprint arXiv:2407.07076 (2024)

[13] ElMahalawy, J., Abbas, O.M., ElSwaify, Y.A., Moustafa, N.: AI powered human computer interaction assisting early identification of emotional and facial symptoms of autism spectrum disorder in children. In: Proceedings of the International Conference on Machine Intelligence and Smart Innovation (ICMISI) (2024)

[14] Rasul, R.A., Saha, P., Bala, D., Karim, S.R.U., Abdullah, M.I., Saha, B.: An evaluation of machine learning approaches for early diagnosis of autism spectrum disorder. arXiv preprint arXiv:2309.11646 (2023)

[15] Farooq, M.S., Tehseen, R., Sabir, M., Atal, Z.: Detection of autism spectrum disorder (ASD) in children and adults using machine learning. Scientific Reports 13(1), 9605 (2023)

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details