Journal of Human Centered Technology

https://www.humentech.utm.my

E-ISSN: 2821-3467

Vision Based Screening of Children with Autism Spectrum Disorders Via Deep Learning Approach

Atiqah Alias[1], Muhammad Amir As'ari[1,2]*

[1]Department of Biomedical Engineering and Health Sciences, Faculty of Electrical Engineering, Universiti Teknologi Malaysia (UTM), 81310 Johor, Malaysia
[2]Sport Innovation and Technology Centre (SITC), Universiti Teknologi Malaysia, Skudai, 81310, Malaysia

*Corresponding Author amir-asari@utm.my

Cite: https://doi.org/10.11113/humentech.v4n2.114

Research Article

Abstract:
Researchers have discerned unique facial characteristics in children with autism, leading to the application of face recognition methods for early diagnosis. Notwithstanding the efficacy of deep learning (DL) models in image classification, issues such as overfitting and limited training datasets continue to exist. This research intends to develop a Convolutional Neural Network (CNN) Deep Learning model for the detection of faces of children with autism. The study employed the Kaggle Autistic Children Facial Dataset to create CNN models through transfer learning, utilizing MobileNet, GoogleNet, and VGG-16. Their performance was assessed using measurements of accuracy, sensitivity, and specificity. The results indicated that MobileNet attained the superior overall performance with accuracy of 0.94, sensitivity of 0.97, and specificity of 0.90, exhibiting robust generalization capabilities. Nonetheless, limits including dataset size and resource limitations were recognized, indicating that future research should prioritize the augmentation of the dataset and the utilization of cloud computing resources for enhanced model training efficiency.

Keywords: ASD; Convolutional Neural Networks; Deep learning; Face recognition

1. INTRODUCTION

Autism is recognized as a multifaceted developmental condition that disrupts the brain's capacity to process information. It emphasizes characteristics such as deficient social skills, absence of eye contact, and repetitive behaviors (1). Social interaction may provide difficulties in sustaining eye contact, comprehending and articulating emotions, and exhibiting suitable facial expressions. Researchers disclosed that children with autism display unique facial characteristics (2). These youngsters typically exhibit a wider upper face and a comparatively shorter midface region. Timely identification is essential for delivering appropriate care and averting developmental complications. Identifying autism is difficult due to its intricacy, resulting in diverse study methodologies, including facial recognition. The emphasis on facial recognition arises from its function in identifying emotional states and differentiating between typical and atypical faces, hence facilitating successful autism detection (1).

Intelligent systems that provide Artificial Intelligence (AI) functionalities typically utilise machine learning (ML) and deep learning (DL). Machine learning is the methodology via which systems acquire knowledge from designated training data pertinent to a certain issue, facilitating the automatic development of analytical models for connected tasks (3). It is classified into supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning (RL). Deep learning (DL) is regarded as a subset of artificial intelligence (AI) and can be perceived as an AI function that emulates the data processing methods of the human brain (4). The actual implementation of machine learning frequently necessitates the integration of deep learning as a fundamental component, employing its technology for tasks such as image or video categorisation, recognition, and identification. Among several deep learning architectures, Convolutional Neural Networks (CNNs) are very effective for picture categorisation, particularly in the early detection of autism spectrum disorder (ASD). Pre-trained CNN models, including VGG-16, GoogLeNet, and AlexNet, are extensively employed in image classification owing to their shown high accuracy. Deep learning can enhance the model's efficacy for a certain job by autonomously extracting image features. As a subset of machine learning, deep learning directly processes raw data by utilising artificial neural networks. Deep neural networks encompass all processing phases, including feature extraction and learning, to enable the development of end-to-end predictive models (5). Deep neural networks are defined by a hierarchical architecture comprising input and output layers, as well as hidden layers that function as representation-learning algorithms. These networks can learn many representations of input data at multiple levels of abstraction. Nonetheless, they are prone to overfitting, especially in scenarios with limited training sets, necessitating datasets containing a considerable quantity of images for efficient training (5).

Received 17 June 2025; Accepted 28 July 2025
Available online 06 August 2025

CNN is recognized as an exceptionally efficient instrument for picture categorization. CNN is a variant of ANN employed for image processing tasks, particularly in computer vision, where it is essential for object recognition in images (5). A Convolutional Neural Network (CNN) consists of an input layer, an output layer, and several hidden layers (6). Each neuron in a specific layer is interconnected with every neuron in the following layer. CNNs can autonomously identify significant features in the initial layers and refine image characteristics in subsequent layers to get precise image categorization. CNN enables automatic feature extraction through its mechanisms. Before that, the creation of an automated facial image categorization system for autistic children utilizing a CNN model could be pivotal in the detection of autism.

Recent studies have focused on the development of automated facial recognition-based diagnostics for numerous conditions, including acromegaly (7), Down syndrome (8), Turner syndrome (9), facial paralysis (10), and autism spectrum disorder (ASD). The research in (11) concentrated on ASD studies and introduced VGG16, VGG19, ResNet18, ResNet101, and DenseNet161 as CNN models for the early detection of ASD utilizing facial image recognition techniques, while the study in (12) compared MobileNetV2 and a hybrid VGG-19 to create a face recognition system leveraging transfer learning. The findings in (12) demonstrate that MobileNet-V2 attains the highest accuracy of 92%. The MobileNet-V1 model has been enhanced, as detailed in (1), to improve the accuracy of detecting children with ASD. Nonetheless, no prior studies have suggested a comparison of the performances of MobileNet, GoogleNet, and VGG-16, on the identical ASD dataset. This study focused on creating an algorithm that identify children with ASD from photographs in the Kaggle dataset, utilizing three proven and robust CNN models: MobileNet, GoogleNet, and VGG-16. The performance of each model was subsequently evaluated.

2. METHODOLOGY

2.1 Main Framework

The software implementation is based on the Python programming language. Figure 1 depicts the comprehensive workflow of the project methodology. The project commences with the collecting and preprocessing of data from the Kaggle Autistic Children Facial Dataset utilized in this endeavour. A CNN model was constructed utilizing transfer learning, followed by the use of various pretraining models to identify autism from face photos. The performance of each training was assessed and validated, encompassing accuracy, sensitivity, and specificity through assessment criteria.

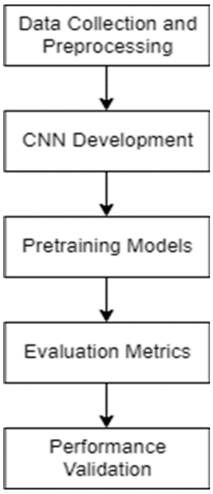

Figure 1. Project's workflow.

2.2 Data Collection and Preprocessing

This study analysed facial images of children with autism and normal children obtained from the publicly accessible Kaggle website. The Kaggle dataset titled Autistic Children Facial Dataset has 2,936 facial images (1468 for autism spectrum disorder and 1468 for normal). Images of children with autism were sourced from websites pertaining to autism spectrum disorder, whilst images of children without autism were randomly obtained from the Internet. The dataset preprocessing, conducted by the dataset creator, involved the elimination of duplicate photos and cropping to focus just on the face. The dataset comprised 1288 photos for training, 140 for validation, and 40 for testing. The dataset has been divided into an 80:20 ratio, with 80% of the photos allocated for training and validation, and 20% designated for evaluating the model's performance, as referenced in (13).

2.3 CNN Development and Pretraining Models

A CNN model for categorising autistic and non-autistic children was created utilizing a transfer learning methodology with pre-trained models, specifically MobileNet, GoogLeNet, and VGG16. The models were trained in the cloud using the Google Colab environment with the Python language, which supports deep learning libraries like TensorFlow. The procedure commenced with the preparation and downloading of the dataset on Google Colab. As previously stated, 80%

https://doi.org/10.11113/humentech.v4n2.114

of the photos were allocated for model training, while the remaining 20% were designated for model testing. Google Drive was implemented to facilitate the storage and retrieval of the dataset, hence enhancing data interaction. This facilitated the straightforward loading of the dataset into the workspace. This stage involved establishing the routes for the training, validation, and test datasets, organizing the dataset into directories with distinct files for each class: 'Autism' and 'Non-Autism'.

Upon loading the photos from the folders, TensorFlow's image_dataset_from_directory function was employed to transform them into datasets. This function autonomously assigns labels to photos according to directory names and randomizes their order to facilitate training. All photos were downsized to a uniform dimension of 224x224 pixels with three colour channels (224x224x3) to ensure uniformity and fulfil model specifications. The pre-trained network was subsequently imported into Google Colab to commence the training process. During this procedure, each model was initialized with weights pre-trained on the ImageNet dataset, omitting the top layers to utilize their feature extraction capabilities. The input photos underwent preprocessing utilizing the preprocess input function specific to each model. This function normalizes the image pixel values to guarantee compatibility with the pre-trained model.

The selected loss function was categorical cross-entropy, and the Adam optimizer operated at a learning rate of 0.00005. A diminished learning rate was employed to guarantee consistent and gradual modifications to the model weights. The cross-entropy loss function assesses the architecture's performance, whereas SoftMax is utilized in the classification layer. Optimizers are algorithms that adjust a neural network's weights and learning rate to minimize losses (14). Numerous types of optimizers are available, including Adagrad, Momentum, Gradient Descent, and RMSProp. This paper employed the Adaptive Moment Estimation (Adam) optimizer. Backpropagation can be executed via mini-batch gradient descent utilizing the Adam optimizer algorithm (15). The models underwent training for 100 epochs, with a batch size of 32, to ascertain the optimal epoch count for maximum testing accuracy. Throughout the training phase, the learning curve was delineated to avert the development of an overfitted model. Among all the models, VGG-16 required the most time to complete training due to its extensive architecture, followed by MobileNet and GoogleNet. The model underwent evaluation following the conclusion of network training.

2.3.1 MobileNet

MobileNet, intended for mobile and embedded vision applications, utilizes depth-wise separable convolutions to create deeper neural networks with a diminished size (16). The MobileNet design employs depth-wise convolution succeeded by pointwise convolution using a 1 x 1 filter to consolidate outputs and reduce model size, hence enhancing its lightweight nature. ReLU and batch normalization are employed in depth-wise separable convolution to mitigate nonlinearity between the two layers. The pre-trained weights from the ImageNet dataset are loaded via keras.applications, the input shape is defined, and the fully connected layer at the apex is omitted. This setup enables the incorporation of an additional output layer for feature extraction and training purposes. Subsequently, the weights are immobilized to retain information across training sessions. A Global Average Pooling layer is incorporated into the original MobileNet model to diminish spatial dimensions and formulate the new model. This option is favoured over Flatten to mitigate excessive parameters and the risk of overfitting. Dense layers are incorporated into the fully connected layers, facilitating the transfer of outputs from preceding layers to the neurons in the current layer, hence enabling the acquisition of intricate functions. Each dense layer is paired with a ReLU activation function that outputs the input value if it is positive; else, it returns zero. The Adam optimizer, a stochastic gradient descent method that incorporates adaptive first- and second-order estimations, is employed to develop the created model. Binary cross-entropy was chosen as the loss function, as it evaluates probabilities against real class values. The input layer of MobileNet has a pixel dimension of 224 × 224 (17).

2.3.2 GoogleNet

GoogleNet is an advanced CNN architecture characterised by a sophisticated design that introduced the "Inception" module. This module efficiently generates a singular filter by amalgamating filters of varying sizes. The GoogleNet architecture comprises nine inception layers, two pooling layers, and two convolutional layers. Each inception layer has six convolutional layers and one pooling layer (18).

2.3.3 VGG-16

The VGG-16 network was trained utilising the ImageNet database. Despite constrained image datasets, the VGG-16 network has exceptional accuracy due to its extensive training. The VGG-16 architecture features a limited receptive field of 3 × 3 and comprises 16 convolutional layers. It contains a maximum pooling layer of dimensions 2 × 2 and comprises a total of five such layers. After the final Max pooling layer, three fully connected layers are present. The final layer employs the softmax classifier. ReLU activation has been applied to each hidden layer (19).

2.4 Performance Evaluation

The metrics specified in Equations 1 to 3 were employed to evaluate the models' efficacy. The classification was based on a confusion matrix that categorised all test dataset images into four groups: True Positive (TP), True Negative (TN), False Negative (FN), and False Positive (FP) (13).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

https://doi.org/10.11113/humentech.v4n2.114

$$Specificity = \frac{TN}{TN + FN} \tag{2}$$

$$Sensitivity = \frac{TP}{TP + FP} \tag{3}$$

A confusion matrix is a tabular representation that evaluates the classification performance of a model, notably in binary classification, such as models used to differentiate between ASD and typically developed (TD). TP denotes instances where the model correctly identified an image as positive or ASD, TN signifies cases where the model accurately classified an image as TD or negative, FP indicates occurrences where the model erroneously categorised a TD image as ASD, and FN refers to situations where the system mistakenly classified an ASD image as TD. This method serves as an essential instrument for evaluating the system's performance and providing insights into strengths and flaws. The combinations of true positives (TP), true negatives (TN), false negatives (FN), and false positives (FP) in the confusion matrix yield performance indicators such as accuracy, sensitivity, and specificity. Accuracy represents the overall percentage of correct predictions made by the model for both positive and negative classifications. Nonetheless, in the presence of an unbalanced dataset, accuracy may be deceptive as it provides a superficial assessment of the model's efficacy. Sensitivity is the measure employed to evaluate a model's predictive capability for true positives within each available category [20]. It assesses the model's capacity to identify positive or target classes. A low false negative rate signifies high sensitivity in the model. Specificity is the measure employed to evaluate a model's ability to accurately predict true negatives for each accessible category (20). The model's specificity indicates its ability to differentiate between negative (or non-target) classes. A high specificity of the model signifies a low false positive rate. These metrics provide a comprehensive assessment of the system's precision in categorizing photos into the correct classifications. Each statistic offers a distinct perspective that assists researchers in evaluating different systems based on their categorizations capabilities and drawing informed judgements.

3. RESULTS AND DISCUSSION

3.1 Confusion Matrix

The performance of MobileNet, GoogleNet, and VGG-16 is illustrated in the confusion matrix utilizing a testing dataset including 40 photos per class, as depicted in Figures 2 to 4. The confusion matrix utilized in this study indicates that the horizontal axis reflects the number of categories for the predicted labels, while the vertical axis denotes the quantity of data corresponding to the genuine labels. This denotes that the values of the matrix at (0,0), (1,0), (0,1), and (1,1) correspond to TP, FP, FN, and TN, respectively. The accuracy rate is determined by dividing the total number of test sets by the sum of the diagonal elements, which signifies the count of model predictions that align with the data labels. In the visualization results, a higher value on the diagonal signifies greater accuracy, while a deeper hue denotes a more precise prediction by the model within that category. Elements off the diagonal signify misclassifications, highlighting regions where the model erred in its predictions.

The confusion matrix of MobileNet demonstrates a highly efficient classification ability, as illustrated in Figure 2. MobileNet accurately identifies a significant quantity of photos in both categories, yielding 39 true positives and 36 true negatives. The FP count of 4 signifies that several non-autistic photos were erroneously categorized as autistic. Conversely, a singular instance of an autistic image was erroneously categorized as non-autistic, as evidenced by the false negative count of one. The results indicate that MobileNet has a relatively low misclassification rate for non-autistic photos while showing exceptional accuracy in recognizing autistic images.

GoogleNet exhibits a somewhat unique performance profile. Figure 3 illustrates that the categorizations proficiency is robust yet not flawless, as evidenced by 35 true positives and 37 true negatives. In comparison to MobileNet, the model demonstrates a higher incidence of 5 false negatives, indicating a greater failure to identify autistic imagery. The FP count of 3, however, is marginally lower than that of MobileNet, suggesting enhanced efficacy in averting the erroneous classification of non-autistic photos. The TP rate is inferior to that of MobileNet, suggesting a greater challenge in reliably identifying all instances of autism, notwithstanding the variations in sensitivity and specificity. The VGG-16 confusion matrix offers a precise assessment of its classification efficacy, as depicted in Figure 4. VGG-16 effectively differentiates between autistic and non-autistic images, achieving 38 true positives and 37 true negatives. The FP count of 3 is analogous to GoogleNet, whilst the FN count of 2 surpasses GoogleNet but is inferior to MobileNet. The results indicate the dependable efficacy of VGG-16, reaching an optimal equilibrium between sensitivity and specificity. VGG-16 is a robust choice for balanced classification performance because to its enhanced ability to reliably identify autistic cases, as seen by its marginally reduced false negative rate in comparison to GoogleNet.

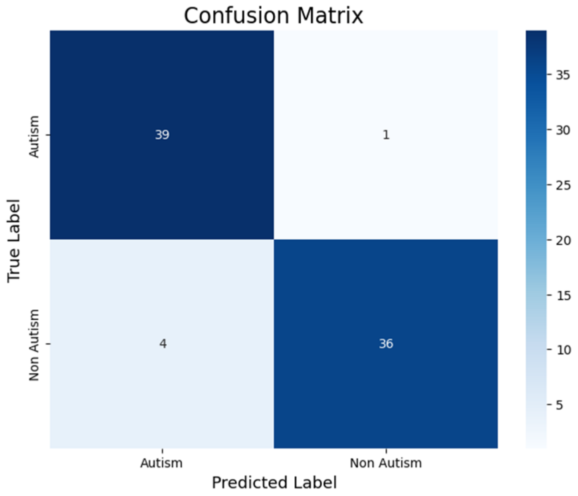

Figure 2. Confusion matrix of MobileNet.

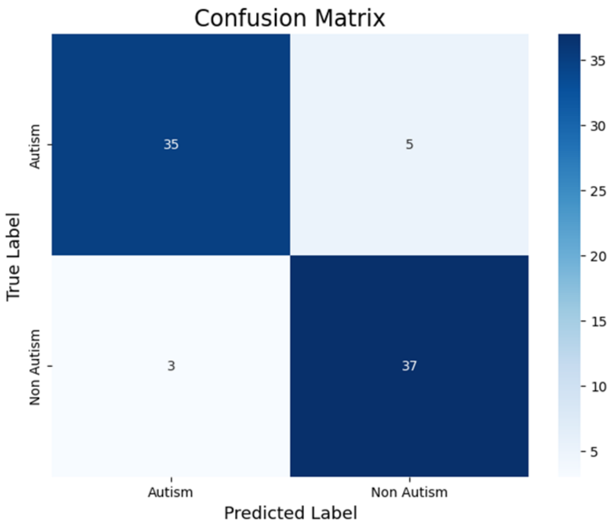

Figure 3. Confusion matrix of GoogleNet.

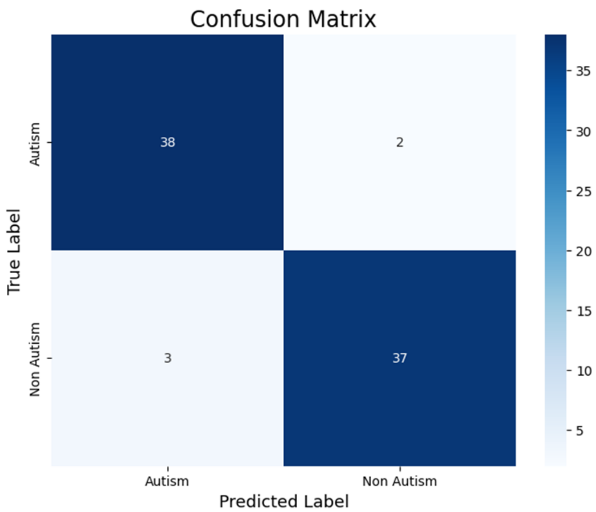

Figure 4. Confusion matrix of VGG-16.

3.2 Performance Metrics of MobileNet, GoogleNet and VGG-16

The three deep learning models, MobileNet, GoogleNet, and VGG-16, were assessed using key performance metrics such as accuracy, sensitivity, and specificity to classify photos into autistic and non-autistic categories over 100 epochs, as presented in Table 1. These metrics offer a thorough insight into the reliability and effectiveness of each model in this essential application.

Table 1. Performance Matrices of MobileNet, GoogleNet and VGG-16 over 100 epochs.

Model	Accuracy	Sensitivity	Specificity
MobileNet	0.94	0.97	0.90
GoogleNet	0.90	0.88	0.93
VGG-16	0.94	0.95	0.93

MobileNet has exceptional efficacy, with an accuracy of 0.94. This signifies that 94% of the photos are accurately categorised by MobileNet. The sensitivity of 0.97 is noteworthy, indicating that it can reliably detect 97% of images related to autism. The elevated sensitivity demonstrates that MobileNet is an invaluable instrument for early diagnosis, signifying its outstanding reliability in identifying cases of autism. Moreover, MobileNet accurately identifies 90% of non-autistic photos, as indicated by its specificity of 0.90, while it exhibits a little higher false positive rate compared to the other models. Comprehensive performance metrics underscore MobileNet's resilience and efficacy in distinguishing between autistic and non-autistic images.

GoogleNet demonstrates marginally inferior metrics compared to MobileNet, however both systems display robust performance. This model correctly detects 90% of the photos, achieving an accuracy of 0.90, which, although inferior to other models, still a respectable rate. With a sensitivity of 0.88, it identifies 88% of photos depicting autism, suggesting a greater proportion of undetected autism cases relative to MobileNet. Nonetheless, GoogleNet offsets this with a specificity of 0.93, signifying its superior ability to reliably differentiate non-autistic images compared to MobileNet. Due to this equilibrium, GoogleNet is a dependable model, particularly in contexts where the reduction of false positives is critical; but its diminished sensitivity may restrict its efficacy in identifying all cases of autism.

VGG-16 achieves an accuracy of 0.94, demonstrating consistent and superior performance across all metrics. The sensitivity of 0.95 signifies a robust ability to accurately identify 95% of autistic cases. This indicates that VGG-16 effectively lowers the number of false negatives by accurately identifying most autistic cases. Moreover, VGG-16 achieves a specificity of 0.93, comparable to GoogleNet, demonstrating its reliability in categorizing photos that are not indicative of autism. VGG-16 is an excellent choice for a balanced classification task due to its equilibrium between sensitivity and specificity, ensuring high detection rates of autistic cases while minimizing false positives.

Comparing these three models demonstrates that MobileNet is the most effective, followed by VGG-16 and GoogleNet, in recognizing all autistic instances and accurately classifying photos into autistic and non-autistic categories, owing to its superior sensitivity of 0.97. This heightened sensitivity is essential for prompt and accurate diagnosis. MobileNet attains a high accuracy of 0.94, equaling VGG-16 and exceeding GoogleNet, so demonstrating its robustness and efficacy in general classification tasks. VGG-16 demonstrates robust performance with an accuracy of 0.94, a sensitivity of 0.95, and a specificity of 0.93, establishing it as a dependable and adaptable choice. GoogleNet demonstrates a sensitivity of 0.88, signifying a higher rate of missed autism cases. Nevertheless, an accuracy of 0.90 and a specificity of 0.93 demonstrate its efficacy in minimizing false positives.

4. CONCLUSION

This work sought to model CNN Deep Learning for the detection of faces of children with autism utilising facial photographs from the Kaggle dataset. The efficacy of pre-trained models MobileNet, GoogleNet, and VGG-16 in categorizing photos into autistic and non-autistic groups was evaluated. This study significantly advances the field of early autism identification via facial recognition by elucidating the strengths and weaknesses of current models.

The findings reveal that the accuracy, sensitivity, and specificity of the three CNN models—MobileNet, GoogleNet, and VGG-16—exhibited robust performance. Of the models evaluated, MobileNet demonstrated the highest efficacy, attaining an accuracy of 0.94, sensitivity of 0.97, and specificity of 0.90. The model's high sensitivity is particularly significant as it underscores MobileNet's reliability in identifying autism cases, which is crucial for quick and accurate identification. VGG-16 exhibited robust performance, effectively balancing sensitivity and specificity for classification tasks. Despite GoogleNet's marginally lower performance, it effectively minimised false positives owing to its superior specificity.

The study concludes that CNN models, especially MobileNet, are highly useful for the early identification of autism via facial recognition. Utilising pre-trained models on extensive datasets can yield high accuracy and reliability in classification jobs. The findings endorse the implementation of these intelligent technologies in clinical environments to facilitate early intervention and enhance outcomes for children with autism. Future research should focus on enhancing the detection skills and resilience of these models by increasing the dataset and investigating various deep learning architectures.

AUTHORSHIP CONTRIBUTION STATEMENT

Atiqah Alias: formal analysis, investigation, writing - original draft; Muhammad Amir As'ari: supervision; validation; writing – review & editing.

DATA AVAILABILITY

https://www.kaggle.com/datasets/imrankhan77/autistic-children-facial-data-set/data

DECLARATION OF COMPETING INTEREST

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

ACKNOWLEDGMENT

The authors would like to express their appreciation to Universiti Teknologi Malaysia (UTM) for endowing this research and the Ministry of Higher Education (MOHE) Fundamental Research Grant Scheme (FRGS/1/2023/ICT02/UTM/02/1) for funding this research.

REFERENCES

(1) Akter T, Ali MH, Khan MI, Satu MS, Uddin MJ, Alyami SA, Ali S, Azad A, Moni MA. Improved transfer-learning-based facial recognition framework to detect autistic children at an early stage. Brain Sci. 2021; 11(6):734. https://doi.org/10.3390/brainsci11060734.

(2) Khosla Y, Ramachandra P, Chaitra N. Detection of autistic individuals using facial images and deep learning. Proceedings of the 2021 IEEE International Conference on Computer Systems and Information Technology for Sustainable Solutions (CSITSS). 2021; 1–5. https://doi.org/10.1109/CSITSS54238.2021.9683205

(3) Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. Electron Markets. 2021; 31(3):685–695. https://doi.org/10.1007/s12525-021-00475-2.

(4) Low WC, Chuah JH, Tee CAT, Anis S, Shoaib MA, Faisal A, Khalil A, Lai KW. An overview of deep learning techniques on chest X-ray and CT scan identification of COVID-19. Comput Math Methods Med. 2021; 2021:1–17. https://doi.org/10.1155/2021/5528144.

(5) Castiglioni I, Rundo L, Codari M, Di Leo G, Salvatore C, Interlenghi M, Gallivanone F, Cozzi A, D'Amico NC, Sardanelli F. AI applications to medical images: From machine learning to deep learning. Phys Med. 2021; 83:9–24. https://doi.org/10.1016/j.ejmp.2021.02.006.

(6) Ahmed ZA, Aldhyani TH, Jadhav ME, Alzahrani MY, Alzahrani ME, Althobaiti MM, Alassery F, Alshaflut A, Alzahrani NM, Al-Madani AM. Facial features detection system to identify children with autism spectrum disorder: Deep learning models. Comput Math Methods Med. 2022; 2022. https://doi.org/10.1155/2022/3941049.

(7) Cakal E. Real-time detection of acromegaly from facial images with artificial intelligence. Eur J Endocrinol. 2023; 188(1):lvad005. https://doi.org/10.1093/ejendo/lvad005.

(8) Qin B, Liang L, Wu J, Quan Q, Wang Z, Li D. Automatic identification of Down syndrome using facial images with deep convolutional neural network. Diagnostics. 2020; 10(7):487. https://doi.org/10.3390/diagnostics10070487.

(9) Pan Z, Zhang D, Zhang Y, Wang M, Yang Y, Xu Y, et al. Clinical application of an automatic facial recognition system based on deep learning for diagnosis of Turner syndrome. Endocrine. 2021; 72:865–873. https://doi.org/10.1007/s12020-020-02539-3.

(10) Amsalam AS, Al-Naji A, Daeef AY, Chahl J. Automatic facial palsy, age and gender detection using a Raspberry PI. BioMedInformatics. 2023; 3(2):455–466. https://doi.org/10.3390/biomedinformatics3020031.

(11) Zheng Y, Liu L. Rapid screening of children with autism spectrum disorders through face image classification. In: Proceedings of the 2022 International Conference on Intelligent Education and Intelligent Research (IEIR); 2022. https://doi.org/10.1109/IEIR56323.2022.10050070.

(12) Alkahtani H, Aldhyani TH, Alzahrani MY. Deep learning algorithms to identify autism spectrum disorder in children-based facial landmarks. Appl Sci. 2023;13(8):4855. https://doi.org/10.3390/app13084855.

(13) Awaji B, Senan EM, Olayah F, Alshari EA, Alsulami M, Abosaq HA, Alqahtani J, Janrao P. Hybrid techniques of facial feature image analysis for early detection of autism spectrum disorder based on combined CNN features. Diagnostics. 2023;13(18):2948. https://doi.org/10.3390/diagnostics13182948.

(14) Tripathi M. Analysis of convolutional neural network-based image classification techniques. J Innov Image Process (JIIP). 2021; 3(2):100–117. https://doi.org/10.36548/jiip.2021.2.003.

(15) Murugan P. Implementation of deep convolutional neural network in multi-class categorical image classification. arXiv preprint. 2018. arXiv:1801.01397. https://doi.org/10.48550/arXiv.1801.01397.

(16) Rahman LA, Booma PM. The early detection of autism within children through facial recognition; a deep transfer learning approach. Proceedings of the 2022 2nd International Conference on New Technologies in Information and Communication (NTIC). 2022; 1–11. https://doi.org/10.1109/NTIC55069.2022.10100517.

(17) Kang J, Gwak J. Ensemble of multi-task deep convolutional neural networks using transfer learning for fruit freshness classification. Multimedia Tools Appl. 2022; 81(16):22355–22377. https://doi.org/10.1007/s11042-021-11282-4.

(18) Hendr A, Ozgunalp U, Erbilek Kaya M. Diagnosis of autism spectrum disorder using convolutional neural networks. Electronics. 2023; 12(3):612. https://doi.org/10.3390/electronics12030612.

(19) Theckedath D, Sedamkar R. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. SN Comput Sci. 2020; 1(2):79. https://doi.org/10.1007/s42979-020-0114-9.

(20) Mitrani A. Evaluating categorical models II: Sensitivity and specificity [Internet]. Medium. 2019 [cited 2023 Dec 6]. Available from: https://medium.com.