https://doi.org/10.31436/iiumej.v25i1.2838

POWER OF ALIGNMENT: EXPLORING THE EFFECT OF FACE ALIGNMENT ON ASD DIAGNOSIS USING FACIAL IMAGES

MOHAMMAD SHAFIUL ALAM[1,2], MUHAMMAD MAHBUBUR RASHID[1*], AHMED RIMAZ FAIZABADI[1] AND HASAN FIRDAUS MOHD ZAKI[1,3]

[1]*Department of Mechatronics Engineering, Kulliyyah of Engineering International Islamic University Malaysia, Kuala Lumpur, Malaysia*
[2]*Department of Electrical and Electronic Engineering, Northern University Bangladesh, Dhaka, Bangladesh*
[3]*Co-ordinator, Center for Unnamed Technology, Kulliyyah of Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia*

[*]*Corresponding author: mahbub@iium.edu.my*

(Received: 15 April 2023; Accepted: 1 August 2023; Published on-line: 1 January 2024)

ABSTRACT: Autism Spectrum Disorder (ASD) is a developmental disorder that impacts social communication and conduct. ASD lacks standard treatment protocols or medication, thus early identification and proper intervention are the most effective procedures to treat this disorder. Artificial intelligence could be a very effective tool to be used in ASD diagnosis as this is free from human bias. This research examines the effect of face alignment for the early diagnosis of Autism Spectrum Disorder (ASD) using facial images with the possibility that face alignment can improve the prediction accuracy of deep learning algorithms. This work uses the SOTA deep learning-based face alignment algorithm MTCNN to preprocess the raw data. In addition, the impacts of facial alignment on ASD diagnosis using facial images are investigated using state-of-the-art CNN backbones such as ResNet50, Xception, and MobileNet. ResNet50V2 achieves the maximum prediction accuracy of 93.97% and AUC of 96.33% with the alignment of training samples, which is a substantial improvement over previous research. This research paves the way for a data-centric approach that can be applied to medical datasets in order to improve the efficacy of deep neural network algorithms used to develop smart medical devices for the benefit of mankind.

ABSTRAK: Gangguan Spektrum Autisme (ASD) adalah gangguan perkembangan yang memberi kesan kepada komunikasi dan tingkah laku sosial. Kelemahan dalam rawatan ASD adalah ianya tidak mempunyai protokol rawatan standard atau ubat. Oleh itu pengenalan awal dan campur tangan betul merupakan prosedur paling berkesan bagi merawat gangguan ini. Kecerdasan buatan boleh menjadi alat berkesan bagi diagnosis ASD kerana bebas campur tangan manusia. Penyelidikan ini mengkaji kesan penjajaran muka bagi diagnosis awal ASD menggunakan imej muka dengan kebarangkalian penjajaran muka dapat meningkatkan ketepatan ramalan algoritma pembelajaran mendalam. Kajian ini menggunakan algoritma penjajaran muka MTCNN berasaskan pembelajaran mendalam SOTA bagi pra-proses data mentah. Selain itu, kesan penjajaran muka pada diagnosis ASD menggunakan imej muka disiasat menggunakan CNN terkini seperti ResNet50, Xception dan MobileNet. ResNet50V2 mencapai ketepatan ramalan maksimum sebanyak 93.97% dan AUC 96.33% dengan sampel penjajaran latihan, yang merupakan peningkatan ketara berbanding penyelidikan terdahulu. Kajian ini membuka jalan bagi pendekatan data berpusat yang boleh digunakan pada set data perubatan bagi meningkatkan keberkesanan algoritma rangkaian saraf mendalam dan membangunkan peranti perubatan pintar bermanfaat untuk manusia.

KEYWORDS: *autism spectrum disorder; CNN; facial images; alignment; deep learning*

1. INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder that can significantly affect a person's life by impairing their ability to communicate, engage with others, and control their repetitive behaviors [1]. There is no precise biomarker to identify ASD, and there is no specific medication to treat the condition. Individuals with ASD can benefit from specific therapy and support services catered to their needs, therefore, early identification and intervention are essential for better outcomes [2].

Machine learning techniques have shown potential for assisting in the early diagnosis and detection of ASD [3]. These techniques make use of algorithms to find patterns and links in massive datasets [4], which can assist in pinpointing important characteristics that set ASD sufferers apart from those without the illness. Working with big datasets of medical pictures or behavioral assessments might greatly benefit from machine learning techniques' capacity to process vast amounts of data rapidly and effectively [5]. The capacity of machine learning techniques to recognize subtle patterns that can escape the attention of human observers is another benefit [6]. For instance, facial recognition algorithms can identify tiny variations in facial traits linked to ASD, even in people who do not show more obvious signs of the condition. Traditionally, diagnosis of ASD has relied on behavioral assessments but conventional interview-based methods have human bias, which leads to the unnecessary delay in detection [7]. Rather recent studies have shown that individuals with ASD exhibit distinct facial features that can be identified using computer vision techniques [8]. However, one major challenge in using facial features for diagnosis is the variability in facial expression and pose across individuals [9]. Face alignment is a technique used to normalize facial images and reduce such variability [10]. In this paper, we investigate the effect of face alignment on ASD diagnosis using facial images and demonstrate the power of alignment in improving the accuracy of ASD diagnosis.

2. METHOD

Autism Spectrum Disorder (ASD) diagnosis is typically a complex and difficult procedure requiring a combination of behavioral assessments, medical evaluations, and standardized tests. Utilizing facial images and analyzing facial features and expressions to aid in the diagnosis of ASD is a promising area of research [11]. However, the accuracy of such methods may be contingent upon the quality and alignment of the facial images employed. Using advanced image processing techniques and machine learning algorithms, we will specifically investigate the effectiveness of alignment in augmenting the recognition of facial features pertinent to ASD diagnosis.

2.1 Dataset

The dataset we are using is the children facial image ASD dataset, which is the only publicly available dataset online containing the facial images of autistic and non-autistic children [12]. This dataset contains a total of 3014 samples of both autistic and non-autistic children of age 2 to 14. The ratio of autistic to non-autistic children is 1:1, while the ratio of male to female children is 3:1. The dataset is divided into three sections: train, test, and validation, with 2654, 280, and 80 samples of ASD and control children, respectively.

Table 1: The formation of ASD dataset

Split	Number	Binary Class
Train set	2654	0.non-ASD 1.Autistic (ASD)
Test set	280	
Valida set	80	

2.2 Dataset Pre-processing

The data has been pre-processed in chronological order. MTCNN, a deep CNN for face detection and alignment, was first used to align the test sample as shown in Fig. 1(a). Three levels of Convolutional Neural Networks in the MTCNN can detect faces and landmarks such eyes, noses, and mouths [13]. After detecting left and right eye coordinates, we can calculate the displacement angle and rotate the picture for the final alignment shown in Fig. 1(a). P-Net, the Proposal Network, is the initial phase of the MTCNN and functions as a proposal generator. Its principal function is to generate prospective face-containing bounding boxes. It consists of three convolutional layers and two fully connected layers. The input filter of P-Net takes the images of 12x12 pixels and the filter sizes in the convolutional layers are 3x3. P-Net generates a list of candidate bounding boxes and their corresponding facial landmark positions. The initial objective of these processes is face recognition, where the cross-entropy loss for each sample is calculated as

$$L_i^{det} = -(y_i^{det} \log (p_i) + (1 - y_i^{det})(1 - \log (p_i))) \tag{1}$$

where p_i is the probability that the sample i = {0,1,...,n} contains a face, as determined by the P-Net, and y_i^{det} is the ground truth level.

The second stage of MTCNN is R-Net, which refines the bounding boxes for different faces, generated by P-Net. There are two fully connected layers followed by three convolutional layers. The input of R-Net is 24x24 pixels, where the filter sizes in convolutional layers are 3x3. R-Net categorizes the candidate boxes as face or non-face and regresses the bounding box coordinates to enhance their precision. For R-Net to construct a bounding box, the four extremities of the box must be located, which is treated as a regression problem, and the Euclidean loss for each sample is computed by multiplying the sample size by the Euclidean loss.

$$L_i^{box} = \| \hat{y}_i^{box} - y_i^{box} \|_2^2 \tag{2}$$

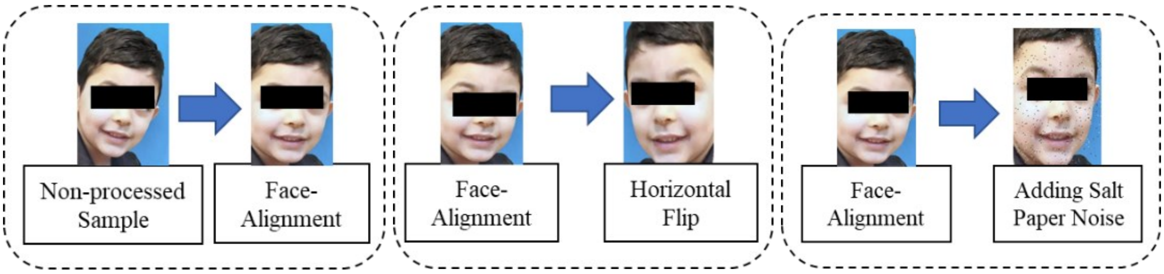
where $\hat{y}_i^{box}$ is the intended level derived from the neural network and y_i^{box} is the coordinate of the ground level.

O-Net, the Output Network, is the last stage of MTCNN and is responsible for the most precise face detection and facial feature alignment. It derives detailed facial features, such as facial landmarks (e.g., eyes, nose, mouth), from the refined candidate boxes provided by R-Net. The construction of O-Net is much more complex than the previous P-Net or R-Net, which consists of three convolutional layers followed by three fully connected. O-Net's input dimension is 48x48 pixels. The dimension of the filters in the convolutional layers is 3x3. For accurate face detection, O-Net classifies the candidate boxes, performs facial landmark localization, and refines the bounding box coordinates. For the creation of the bounding box, four coordinates such as top, breadth, and height are necessary, thus $y_i^{box} \in \mathbb{R}^4$. In the final phase, the Euclidean loss is again minimised according to the following equation to formulate the face landmark detection task.

$$L_i^{landmark} = \| \hat{y}_i^{landmark} - y_i^{landmark} \|_2^2 \tag{3}$$

where $\hat{y}_i^{landmark}$ is the co-ordinates of facial landmarks - Left eye, right eye, nose, left corner of mouth and right corner of mouth and $y_i^{landmark}$is the ground truth co-ordinate for the *i*th number of images, thus $y_i^{landmark} \in \mathbb{R}^{10}$.After the detection of the left and right eye co-ordinates we can get the angle θ and the image has to be rotated anti-clockwise at an angle θ for Alignment.

Figure 1(b) depicts the subsequent phase of the pre-processing, which is the horizontal flip of the test samples. This action is performed after the alignment procedure and is also used as a separate training set for deep learning models. The horizontal flip improves the training of deep learning models because it increases the quantity of training data and provides the model with more diverse examples from which to learn [14]. A second training set is created by combining the facial image samples with the basic Gaussian salt and pepper noise, as depicted in Fig. 1(c). Previous research predicted that low image quality would negatively impact the training and efficacy of the deep learning model [15]. Adding noise provides the algorithm with a chance to deal with low-quality training samples, which may slightly boost its performance.



(a) (b) (c)

Fig. 1: (a) Face Alignment (b) Horizontal flip and (c) Noise addition pipeline for training samples.

2.3 Experimental Setup

The model is trained on Kaggle platform using the Tensorflow library. In this experiment, deep learning models ResNet50V2 [16], Xception [17], and MobileNetV2 [18] are selected based on the high accuracy reported by Alam et al. [19] with transfer learning approach as shown in Fig. 2.

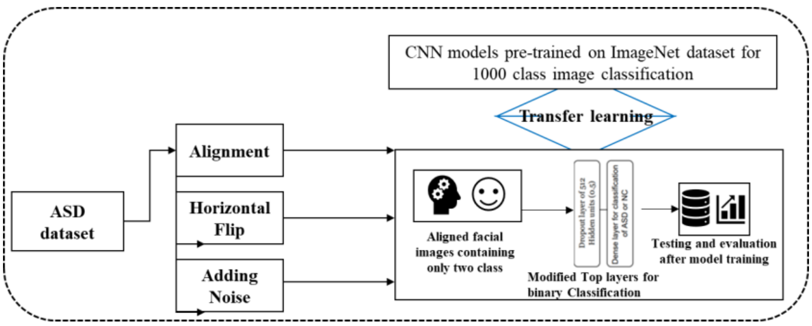


Fig. 2: ASD diagnosis process using facial image and transfer learning approaches.

All these CNN models are pretrained on the large-scale ImageNet database using a supercomputer to determine the initial weights for image classification tasks. The hyperparameters for model training are held constant in accordance with the best configuration from the same model-centric research. Adagrade is chosen as the optimizer, the batch size is maintained at 32, and the learning rate is set to 0.001. For evaluating performance, Accuracy, AUC, Precision, and Recall were considered as performance metrics to benchmark with existing methods.

3. RESULTS AND DISCUSSION

3.1 Performance Evaluation for Different Training Sets

In this section, the results obtained from the distinct training sets were analyzed and presented after face alignment and further processing. Table 2 shows the detection performance of various CNN models with face alignment. The validation and test sets are left unaltered in order to compare the results with recent research. After applying flip and noise addition separately to the aligned training set, the performance evaluation is done again on the test set. Finally, all three sets are combined for training resulting in three times as many training samples as the original training set and later is evaluated with same test set. For face alignment only, ResNet50V2 achieves the highest test accuracy of 93.93% with 96.33% AUC, while the Xception model also performs very closely with 92.14% accuracy and 95.91% AUC. After applying flip and noise to the aligned dataset during training, the test results are not as impressive; rather, the prediction accuracy drops dramatically. The test accuracy is recorded at 88.57% with 93.30% AUC for the ResNet50V2 model applying flip, but the results of other training sets with other models are not good enough to be considered. By combining all three sets while training three distinct deep CNN models, ResNet50V2 achieved the highest performance having 92.86% accuracy and 95.65% AUC which is actually lower than that accuracy of applying face alignment only.

Table 2: The performance of different deep learning models with face alignment

Dataset Processing	Training Dataset size	Algorithm	Accuracy	AUC	Precision	Recall	Loss
Face Alignment	2654	ResNet50V2	**93.93**	**96.33**	**93.93**	**93.93**	**0.373**
		Xception	92.14	95.91	92.14	92.14	0.360
		MobileNetV2	84.64	92.20	84.64	84.64	0.451
Face Alignment with Flip	2654	ResNet50V2	88.57	93.30	88.57	88.57	0.576
		Xception	87.50	94.06	87.50	87.50	0.482
		MobileNetV2	82.14	87.23	82.14	82.14	0.742
Face Alignment with Noise Addition	2654	ResNet50V2	72.50	80.66	72.50	72.50	1.017
		Xception	88.21	93.91	88.21	88.21	0.351
		MobileNetV2	63.21	72.33	63.21	63.21	1.247
Face Alignment+ Flip+Noise Addition	7062	ResNet50V2	92.86	95.65	92.86	92.86	0.496
		Xception	92.50	94.75	92.50	92.50	0.473
		MobileNetV2	86.43	91.63	86.43	86.43	0.718

Table 3 displays the training accuracy, T_{acc}, and validation accuracy, V_{acc}, for various deep CNN models while training with distinct datasets produced by dataset preprocessing. Face alignment for ResNet50V2 has a T_{acc} of 99.5%, plainly demonstrating the justification for performing it during training. After employing flip augmentation and merging all processed datasets for training, T_{acc} for ResNet50V2 is increased to nearly 99.8%. Using all three training sets (Face Alignment + Flip+ Noise Addition) cumulatively for training, the Xception model also demonstrates a very good training performance with an accuracy of 99.7%. Increase in training accuracy indicates that models are learning more effectively from the training set and are expected to perform well during testing. The 99.8% validation accuracy for face alignment demonstrates the superior learning capability of ResNet50V2, while the models' performances are sufficiently good enough to contribute the value for this research work.

Table 3: The Training (T_{acc}) and validation accuracy (V_{acc}) after Face alignment and applying augmentations

	Face Alignment		**Face Alignment with Flip**		**Face Alignment with Noise Addition**		**Face Alignment + Flip+ Noise Addition**	
	T_{acc}	V_{acc}	T_{acc}	V_{acc}	T_{acc}	V_{acc}	T_{acc}	V_{acc}
ResNet50V2	0.995	**0.988**	**0.998**	0.950	0.995	0.775	**0.998**	0.975
Xception	0.988	0.975	0.991	0.950	0.987	0.938	0.997	0.975
MobileNetV2	0.971	0.938	0.970	0.888	0.965	0.688	0.989	0.950

Figure 3 depicts the aforementioned scenario with graphical representations, demonstrating that ResNet50V2 outperforms all other models for aligned facial images of ASD and control children. According to the graph, the performance is lowest for noise addition with aligned training set, which has a testing accuracy of only 62.31%.

Figure 4 shows the training and validation accuracy curve for the best result after face alignment of the training samples. The training is consistent with the validation curve, and no significant overfitting is observed, as the validation accuracy follows the training curve until its convergence. The trend holds true for validation loss as well. Figure 5 depicts the confusion matrix for the identical experimental configuration with face alignment on training samples. The number of missed predictions is indicated by the white boxes, and for the ResNet50V2 model, only one control child was incorrectly classified as autistic, while the number of wrong predicted images for the autistic sample is 16. Thus overall, the number total is 17 while the numbers for missed predictions are 22 and 43 for Xception and MobileNetV2 respectively which makes the ResNet50V2 superior to other models for face alignment on training set.

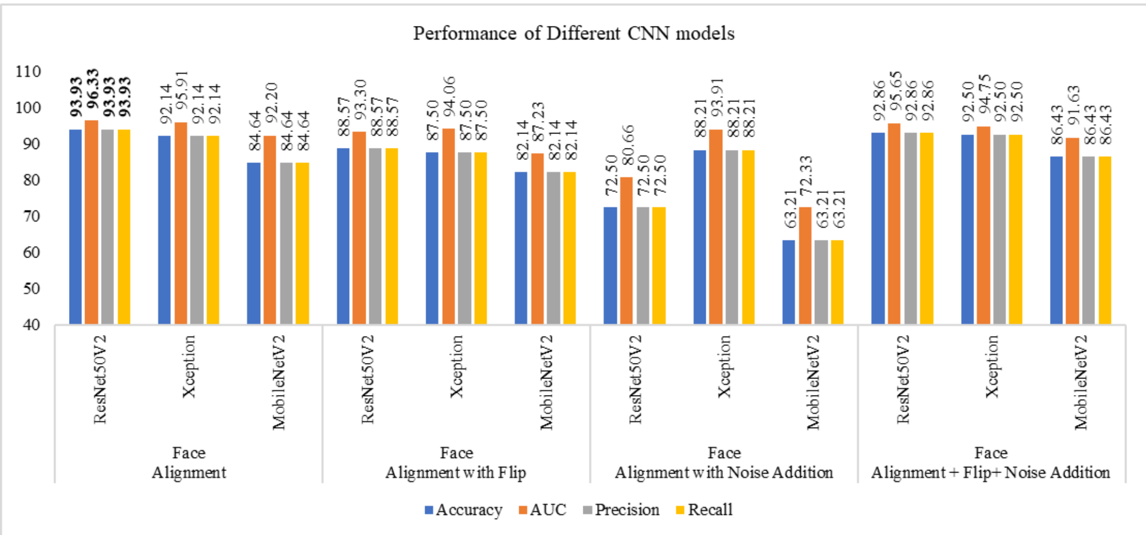
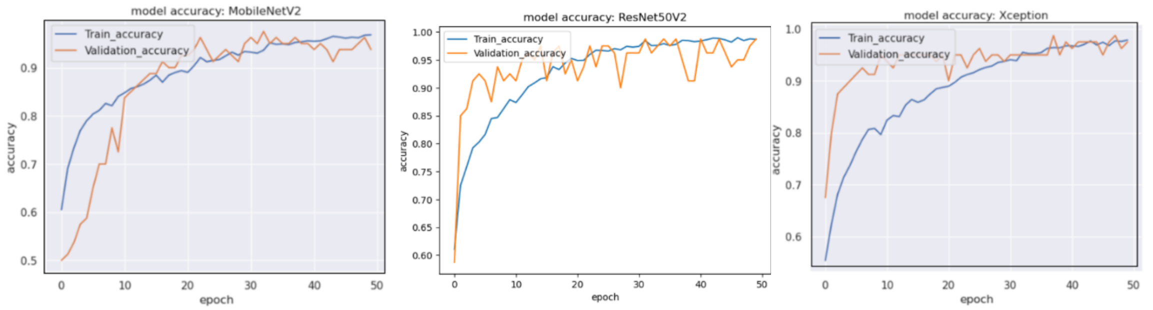
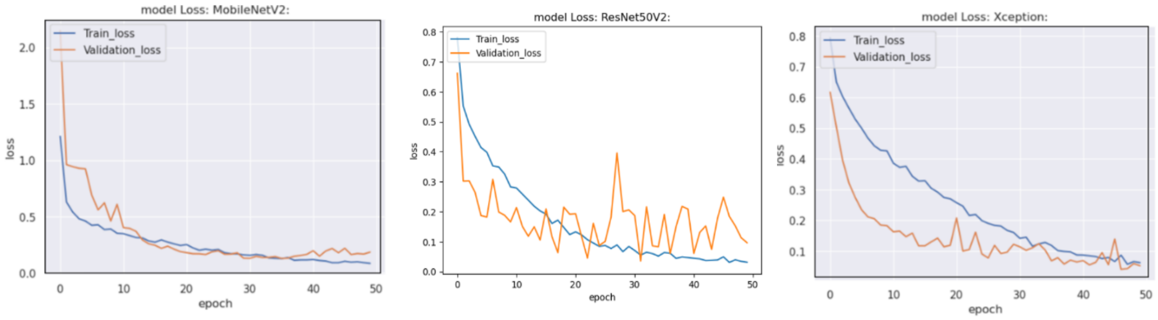


Fig. 3: Performance graph for the different CNN models for a specific training set.

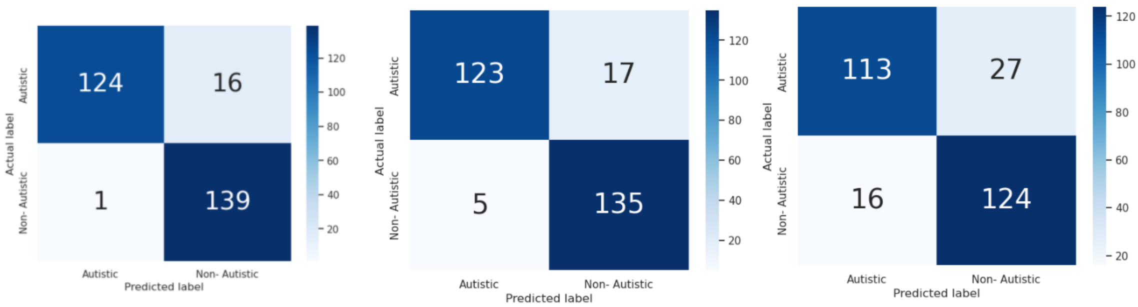


(a) (b) (c)



(a) (b) (c)

Fig. 4: Graphical representation of training and validation accuracy of (a) ResNet50V2, (b) MobileNetV2 and (c) Xception model and training and validation loss of (a) ResNet50V2, (b) MobileNetV2 and (c) Xception model for face alignment.



(a) (b) (c)

Fig. 5: Confusion matrix of (a) ResNet50V2, (b) Xception and (c) MobileNetV2 model for the face alignment.

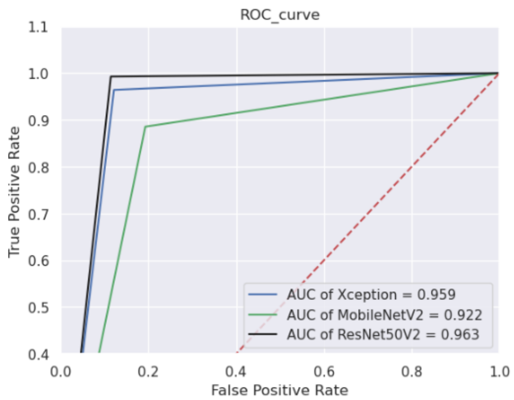


Fig. 6: ROC curve of (a) ResNet50V2, (b) Xception and (c) MobileNetV2 model for the face alignment.

ROC displays the area under the curve, AUC, which provides the region that states the model's coverage; the larger the area, the greater the model's ability to detect the correct class without any error. Figure 6 displays the ROC for three models with values for the best performing experimental setup - alignment of facial images of the training set. Based on the ROC plot, ResNet50V2 has a larger area, indicating a greater likelihood of correctly detecting the sample.

3.2 Discussion

In this study the effect of face alignment on the diagnosis of Autism Spectrum Disorder (ASD) is explored using facial images. Here we used ASD dataset containing facial images of children with ASD and typically developing children. The dataset contained facial image samples which were fed through MTCNN for face alignment. It is hypothesized that aligned facial images can improve the accuracy of ASD diagnosis by reducing variations in head position and facial expressions.

The diagnosis of ASD using facial images and a deep learning algorithm is a relatively novel field of research. ASD is a disorder characterized by a deficiency in neurological development and the human face can reveal information about brain function and structure [20]; therefore, the face could be a crucial biomarker for diagnosing ASD. Using deep learning, researchers are attempting to identify facial characteristics that correlate with neurological disorders which are invisible to the human eye. Table 4 shows a very brief comparison of the performance metrices of the contemporary research works. In early 2021, Rabbi et al. [21] accomplished a very good prediction accuracy by training their own CNN using ASD facial image dataset. Later, Arumugam et al. [22], Shaik et al. [23] and Kaur et al. [24] obtained 91%, 84.67% and 70% respectively using VGG16. Akter et al. [25] has performed training using shallow and deep methods and achieved an accuracy of 92.1% using MobileNet-V1. Further, Rahman et al. [15], Alsaade et al. [26] obtained 90~91 percent prediction accuracy using Xception. Recently, Alam et al. [19] performed an ablation study and achieved the highest accuracy (95%) of all.

Table 4: Comparison of performance parameters with the recent research

Ref	CNN Model	Sample size	Accuracy %	Precision	Recall	Data pre-processing
Rabbi et al. [21]	own CNN	2940	92.32	89.72	93.45	None
Arumugam et al. [22]	VGG16	2940	91.00	-	-	None
Shaik et al. [23]	VGG16	2940	84.67	-	-	None
Akter et al. [25]	MobileNet-V1	2940	92.10	92.10	92.10	None
Rahman et al. [15]	Xception	2940	90.00	-	-	None
Alsaade et al. [26]	Xception	2940	91.00	-	-	None
Alam et al. [19]	Xception	3014	95.00	95.00	95.00	Cleaning
Kaur et al. [24]	VGG16	2940	70.00			None
Our Proposed Face Alignment						
Face Alignment	ResNet50V2	3014	**93.93**	93.93	93.93	Alignment
Face Alignment with Flip	ResNet50V2	3014	88.57	88.57	88.57	Flip
Face Alignment with Noise Addition	ResNet50V2	3014	72.50	72.50	72.50	Noise Addition
Face Alignment + Flip+ Noise Addition	ResNet50V2	8322	92.86	92.86	92.86	All

Dataset pre-processing techniques such as Align, Horizontal flip, or noise addition were not being used by any researchers prior to this work to improve the quality and increase the number of samples of the training data. Image alignment is very necessary to bring symmetry to the dataset by aligning the facial landmarks that result in better outcomes in classification [27]. After aligning the dataset, the highest reported prediction accuracy in this study is 93.93 percent. Despite the fact that employing other augmentations after image alignment improves training accuracy according to Table 2, the evaluation of the test dataset is not remarkable. The training evaluation performance of the synthesized training dataset, comprising all three combinations, is very promising since evaluating with the test dataset results in lower values of accuracy than that of using only face alignment.

Data augmentation introduces high variance in the samples of the training dataset, which can negatively impact evaluation performance. The variations introduced to both classes bring about certain similarities which cause overfitting, limited generalization, increased misclassifications and lastly, impose increased computational requirements. Relatively applying face alignment improved the detection accuracy by outperforming almost all the previous research, as shown in Table 3. Our research demonstrates a lower value than the 95% accuracy reported by Alam et al. [19], suggesting that alignment alone is insufficient to improve the performance of deep learning algorithms. Rather, a detailed data-centric strategy should be investigated in greater depth. Poor image quality and inadequate medical validation of the ASD patients present in the training set are additional limitations of this study.

4. CONCLUSION

Our study highlights the importance of face alignment in improving the accuracy of Autism Spectrum Disorder (ASD) diagnosis using facial images. We explored the effect of alignment on the classification performance of three state-of-the-art deep learning models, and our results showed a significant improvement in prediction accuracy when using aligned faces compared to unaligned faces. Our findings suggest that accurate face alignment can improve the quality of facial features used by machine learning models, leading to more accurate and reliable predictions. The best performance was achieved with alignment only, where we achieved a prediction accuracy of 93.97% with 96.33% AUC using ResNet50V2. This is a significant improvement over previous research which has important implications for the diagnosis of ASD, where accurate and early diagnosis is crucial for effective intervention and treatment. We believe that our study opens up new avenues for future research on the use of face alignment for ASD diagnosis using facial images, and we hope that our results will contribute to the development of more innovative tools for medical professionals.

ACKNOWLEDGEMENT

The authors would like to express the deepest gratitude to the International Islamic University Malaysia for its support through the Tuition Fee Waiver Scheme (2021) to M.S.A.

REFERENCES

[1] Ghosh T, Banna MHA, Rahman MS, Kaiser MS, Mahmud M, Hosen ASMS, Cho GH. (2021) Artificial intelligence and internet of things in screening and management of autism spectrum disorder. Sustainable Cities and Society, 74: 103-189. https://doi.org/10.1016/j.scs.2021.103189

[2] Al Banna MH, Ghosh T, Taher KA, Kaiser MS, Mahmud M. (2020) A Monitoring System for Patients of Autism Spectrum Disorder Using Artificial Intelligence. The 13th International Conf. on Brain Informatics (BI 2020), 251-262.https://doi.org/10.1007/978-3-030-59277-6_23

[3] Cao, X., & Cao, J. (2023). Commentary: Machine learning for autism spectrum disorder diagnosis--challenges and opportunities--a commentary on Schulte-Rüther et al. (2022). Journal of Child Psychology and Psychiatry, 64(6):966–967. https://doi.org/10.1111/jcpp.13764

[4] Islam, M. R., Rashid, M. M., Rahman, M. A., Mohamad, M. H. S. Bin, & others. (2022). Analysis of blockchain-based Ripple and SWIFT. Asian Journal of Electrical and Electronic Engineering, 2(1):1–8. https://alambiblio.com/ojs/index.php/ajoeee /article/view/26

[5] Bokshi, L. R., Al Banna, M. H., Ghosh, T., Al Nahian, M. J., & Kaiser, M. S. (2022). Investigation on Heart Attack Prediction Based on the Different Machine Learning Approaches. Rhythms in Healthcare, 95–108. Springer. https://doi.org/10.1007/978-981-19-4189-4_7

[6] Ghosh, T., Banna, M. H. A., Nahian, M. J. A., Kaiser, M. S., Mahmud, M., Li, S., & Pillay, N. (2022, September). A privacy-preserving federated-mobilenet for facial expression detection from images. In International Conf on Applied Intelligence and Informatics, 277-292. https://doi.org/10.1007/978-3-031-24801-6_20

[7] Landa, R. J. (2008). Diagnosis of autism spectrum disorders in the first 3 years of life. Nature Clinical Practice Neurology, 4(3):138–147. https://doi.org/10.1038/ncpneuro0731

[8] Ahmed, Z. A. T., Aldhyani, T. H. H., Jadhav, M. E., Alzahrani, M. Y., Alzahrani, M. E., Althobaiti, M. M., Alassery, F., Alshaflut, A., Alzahrani, N. M., & Al-madani, A. M. (2022). Facial Features Detection System To Identify Children With Autism Spectrum Disorder: Deep Learning Models. Computational and Mathematical Methods in Medicine, 2023:1–9. https://doi.org/10.1155/2022/3941049

[9] Cowen, A. S., Keltner, D., Schroff, F., Jou, B., Adam, H., & Prasad, G. (2021). Sixteen facial expressions occur in similar contexts worldwide. Nature, 589(7841):251–257. https://doi.org/10.1038/s41586-020-3037-7

[10] Abdollahi, B., Tomita, N., & Hassanpour, S. (2020). Data augmentation in training deep learning models for medical image analysis. Deep Learners and Deep Learner Descriptors for Medical Applications, 186:167–180. https://doi.org/10.1007/978-3-030-42750-4_6

[11] Briot, K., Pizano, A., Bouvard, M., & Amestoy, A. (2021). New Technologies as Promising Tools for Assessing Facial Emotion Expressions Impairments in ASD: A Systematic Review. Frontiers in Psychiatry, 12. 634756 https://doi.org/10.3389/fpsyt.2021.634756

[12] Musser, M. (2020). Detecting Autism Spectrum Disorder in Children With Computer Vision. Towards Data Science. Available: https://github.com/mm909/Kaggle-Autism

[13] Zhang, N., Luo, J., & Gao, W. (2020). Research on Face Detection Technology Based on MTCNN. 2020 International Conference on Computer Network, Electronic and Automation (ICCNEA), 2020:154–158. https://doi.org/10.1109/ICCNEA50255.2020.00040

[14] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1):1-48 https://doi.org/10.1186/s40537-019-0197-0

[15] Mujeeb Rahman, K. K., & Subashini, M. M. (2022). Identification of Autism in Children Using Static Facial Features and Deep Neural Networks. Brain Sciences, 12(1):94. https://doi.org/10.3390/brainsci12010094

[16] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity mappings in deep residual networks. In Computer Vision–ECCV 2016: 14th European Conf Proceedings, 14:630-645. https://doi.org/10.1007/978-3-319-46493-0_38

[17] Chollet, F. (2016). Xception: Deep Learning with Depthwise Separable Convolutions. IEEE Conf. on Computer Vision and Pattern Recognition,1251–1258. http://arxiv.org/abs/1610.02357

[18] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. IEEE Conf. on Computer Vision and Pattern Recognition, 4510–4520. https://doi.org/10.48550/arXiv.1801.04381

[19] Alam, M. S., Rashid, M. M., Roy, R., Faizabadi, A. R., Gupta, K. D., & Ahsan, M. M. (2022). Empirical Study of Autism Spectrum Disorder Diagnosis Using Facial Images by Improved Transfer Learning Approach. Bioengineering, 9(11):1–18. https://doi.org/10.3390/bioengineering9110710

[20] Porto, J. A., Bick, J., Perdue, K. L., Richards, J. E., Nunes, M. L., & Nelson, C. A. (2020). The influence of maternal anxiety and depression symptoms on fNIRS brain responses to emotional faces in 5-and 7-month-old infants. Infant Behavior and Development, 59:101447. https://doi.org/10.1016/j.infbeh.2020.101447

[21] Rabbi, M. F., Hasan, S. M. M., Champa, A. I., & Zaman, M. A. (2021). A Convolutional Neural Network Model for Early-Stage Detection of Autism Spectrum Disorder. 2021 International Conf. on Information and Communication Technology for Sustainable Development, ICICT4SD – 2021:110–114. https://doi.org/10.1109/ICICT4SD50815.2021.9397020

[22] Arumugam, S. R., Karuppasamy, S. G., Gowr, S., Manoj, O., & Kalaivani, K. (2021). A Deep Convolutional Neural Network based Detection System for Autism Spectrum Disorder in Facial images. Proceedings of the 5th International Conf on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2021, 1255–1259. https://doi.org/10.1109/I-SMAC52330.2021.9641046

[23] Jahanara, S. (2021). Detecting autism from facial image. International Journal of Advance Research, Ideas and Innovations in Technology, 7(2):219-225. https://doi.org/10.13140/RG.2.2.35268.35202

[24] Kaur, N., & Gupta, G. (2023). Refurbished and improvised model using convolution network for autism disorder detection in facial images. Indonesian Journal of Electrical Engineering and Computer Science, 29(2):883–889. https://doi.org/10.11591/ijeecs.v29.i2.pp883-889

[25] Akter, T., Ali, M. H., Khan, M. I., Satu, M. S., Uddin, M. J., Alyami, S. A., Ali, S., Azad, A., & Moni, M. A. (2021). Improved Transfer-Learning-Based Facial Recognition Framework to Detect Autistic Children at an Early Stage. Brain Sciences, 11(6):734. https://doi.org/10.3390/brainsci11060734

[26] Alsaade, F. W., & Alzahrani, M. S. (2022). Classification and Detection of Autism Spectrum Disorder Based on Deep Learning Algorithms. Computational Intelligence and Neuroscience, 2022:1–10. https://doi.org/10.1155/2022/8709145

[27] Fu, Y., Lei, Y., Wang, T., Curran, W. J., Liu, T., & Yang, X. (2020). Deep learning in medical image registration: a review. Physics in Medicine & Biology, 65(20), 1-50. https://doi.org/10.1088/1361-6560/ab843e