UTM UNIVERSITI TEKNOLOGI MALAYSIA
INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING
ISSN 2180-4370
Journal Homepage : https://ijic.utm.my/

Facial Expression Recognition of Autism Individuals for Analyzing Emotions Using Machine Learning

Lee Qing Ren[1] & Pang Yee Yong[2]
Faculty of Computing
Universiti Teknologi Malaysia
81310, Johor Bahru, Johor, Malaysia
Email: leeren@graduate.utm.my[1], yeeyongpang@utm.my[2]

***Abstract*—Autism individuals have the difficulties in expressing facial emotion. Instead, they have abnormal facial expressions to express their emotions which are hardly recognizable. Meanwhile, the computational cost increases along with the size of dataset. Therefore, this research aims to propose a facial emotion recognition model designated for autistic people and focus on investigating approaches that use lesser computational time to extract features in an image. This research proposed to use the least computational time feature extraction approach, which is Principal Component Analysis (PCA), and the fully connected layer of neural network for the autistic emotion classification. The dataset used is small with 782 images for 4 emotions, such as anger, joy, natural and sadness. The images were preprocessed and normalized before undergoing each feature extraction approach. To measure the performance of the feature extraction approaches such as PCA, Linear Discriminant Analysis (LDA), PCA+LDA and facial landmarks localization, computational time and dimensionality shape of the extracted data were investigated. The least computational time feature extraction approach, PCA was used to extract the features for the classifiers, such as the proposed model, Multilayer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, Decision Tree and AdaBoost. Based on the results, the proposed model had the highest classification accuracy among the classifiers (64.30%). Joy was well predicted than the others with higher values in the confusion matrix.**

Keywords — Facial Emotion Recognition, Autism Spectrum Disorder, Machine Learning, Feature Extraction, Computational Time.

I. INTRODUCTION

Individuals with Autism Spectrum Disorder (ASD) is lack of expressing meaningful facial expression to express their genuine emotion due to the developing brain is affected by both genetic and environmental factors [1], [2], [3]. This becomes a challenge in clinical diagnosis and treatment as doctors are unable to comprehend the patients' thoughts and emotions [4].

Machine learning is introduced in many fields, one of the fields is image classification. In this field, the well-trained model can classify emotions accurately from image inputs. Although the model can reach high accuracy in classifying emotions of neurotypical people, it is a challenge to reach high accuracy in classifying emotions of autistic people as the facial expressions they made are abnormal [4]. It is hard to define a pattern to classify their emotions humanly, such as joy, sadness, fear, natural, anger and surprise.

As machine learning becomes more popular, there is a demand to increase efficiency of model training [5]. This is because the computational cost increases with the dataset size [5]. Training a model with a huge dataset requires many GPU resources to support.

Therefore, in this research, we propose a model that applies less training time in feature extraction to recognize the autistic people's emotions while maintaining the classification accuracy as well.

II. RELATED WORKS

Autism Spectrum Disorder, which is ASD in short, refers to the individuals who are often living in their own world, paying no attention to the surroundings [6]. Many studies justified that one of the criteria of ASD people is reduced sharing of feelings and emotions [1], [2], [7]. In other words, it means that the way of ASD people expressing their emotions through their facial expressions are different from the normal people, and this is one of the reasons why their social interactions are not effective as it is difficult to interpret the abnormal facial emotions [4]. Therefore, it is a huge challenge for ASD medical treatment and parental care.

There were research trying to come out with a reliable facial emotion recognition model to predict the emotions from facial expressions. Based on the findings of Michel and Kaliouby [7], the emotion recognition accuracy of the proposed real time facial expression system using Support Vector Machine (SVM) was 86% and 71.8% using still image and video respectively. Different emotions had different classification accuracy, and this was because of the head position and motion [7].

A comparison between the autistic children's emotion classification accuracy of SVM and Artificial Neural Network (ANN), which is also known as Multilayer Perceptron (MLP), was carried out by Rani [8], and the result showed that SVM could achieve higher accuracy than ANN.

Meanwhile, decision tree is also one of the options of the emotion classifier. Lee et al. [9] introduced a hierarchical binary decision tree approach to make emotion classification using the signal data of the spoken utterance and tone. The approach aimed to minimize the error propagation by having the classifier to predict the easiest emotions.

Random forest, which is also a tree based like decision tree, was integrated with Convolutional Neural Network (CNN) by inserting it at the last pooling layers of CNN, to predict 4 different facial emotions and to overcome the computational problem, especially in the feature extraction stage [10].

Owusu et al. [11] used Adaptive Boosting, which is also called AdaBoost in short, in the proposed model to find out the higher emotion classification accuracy in JAFFE and YALE datasets and the result showed that the emotion classification accuracy of the proposed model using JAFFE dataset was higher than that of using YALE dataset.

Dino and Abdulrazzaq [12] made a comparison between some classifiers to identify which classifier could achieve the best emotion classification accuracy while using less feature inputs and the result showed that KNN was the classifier with only 30 features inputs.

Mollahosseini et al. [13] proposed a new CNN based neural network by adding 4 inception layers and increasing the number of neurons and layers to improve the neural network capability in the facial expression classification.

Silva et al. [14] customized CNN to recognize the autistic facial emotions in real time from those who might have a meltdown with 80.60% accuracy with Face Expression Recognition Plus (FER+) dataset. The feature extraction that Silva et al. [14] utilized was PCA to extract the important features of the facial expressions of autistic children.

Based on the previous work, research and studies that focused on autistic individuals are less. Most of the comparison studies and proposed models did not address the distinct facial expression characteristics of autism people specifically. Therefore, there is a need to come out with a model specifically tailored to recognizing the facial emotions of the people with autism since they have different ways to express their facial emotions compared to the normal people.

For the related works in feature extraction, Marasamy and Sumathi [15] applied wavelet transforms to improve the performance of the wavelet Linear Discriminant Analysis (LDA) in facial recognition task. In terms of time, PCA and LDA reached the optimized accuracy faster than the wavelet LDA when the number of data fit increased [15].

An analysis of the combination of PCA and LDA was carried out to find this combination performance in facial expression recognition using SVM and HMM classifier and the result showed that the combination performs better in SVM [16].

In the research of Sun et al. [17], it was found that facial landmarks localization performed better than LDA as the classification accuracy of emotion recognition was higher.

These related works about the feature extraction approaches focused on improving accuracy which would increase the computational complexity in return. There is a need to address the problem of computational time while maintaining the optimized model accuracy.

III. Proposed Methods

This section will describe the proposed methods used.

A. Datasets

The dataset used is combined 2 autistic facial emotion datasets, one dataset was obtained from the website Kaggle with the URL: https://www.kaggle.com/datasets/fatmamtalaat/autistic-children-emotions-dr-fatma-m-talaat provided by Talaat [18] and another was obtained from Mashuque Alamgir et al. [19] in their introduced novel of facial expressions database of autistic children.

Fig. 1. Autistic Facial Expressions

Fig. 1 shows the examples of the autistic facial expressions. The emotion classes such as fear and surprise are dropped as the difference between the number of images in these emotion classes and the others is significant, which will result in imbalanced data, biased classification and overfitting. Hence, there are currently 4 emotion classes which have 782 images in total, such as 100 angry, 350 happy, 233 sad and 99 neutral. The dataset is split into 80% training and 20% validation.

B. Preprocessing

The images for each class will be converted into grayscale iteratively. Then the images undergo face detection to detect face and crop the face detected, assuming a face in each image. It is then resized into (128, 128) before data augmentation. Lastly, normalization is carried out on the data. However, different approaches will have different methods in normalizing the data.

Standard Scaler is used in normalization for LDA, PCA and PCA+LDA while histogram equalizer is used in normalization for facial landmarks localization.

C. Feature Extraction

There are 4 approaches to compare to find out the least computational time approach such as PCA, LDA, PCA+LDA and facial landmarks localization. The performance metrics to evaluate the approaches are the execution time and dimensionality shape reduced. PCA, LDA and PCA+LDA will be implemented using the built-in functions provided by sklearn while facial landmarks localization will be implemented using dlib library.

D. Classifiers

There are a total of 7 classifiers to be compared in terms of autistic facial emotion classification accuracy, such as the proposed model, SVM, MLP, KNN, Random Forest, Decision Tree and AdaBoost. The classifiers are initialized to their suitable values in the parameters to achieve the optimized result in accuracy. Meanwhile, the proposed model is a customized neural network, so the parameters are set to epoch=50, shuffle=True and batch size=32. All classifiers use the extracted features from the feature extraction approach with the least computational time. The performance metric to evaluate the classifiers is accuracy.

IV. Experimental Setup

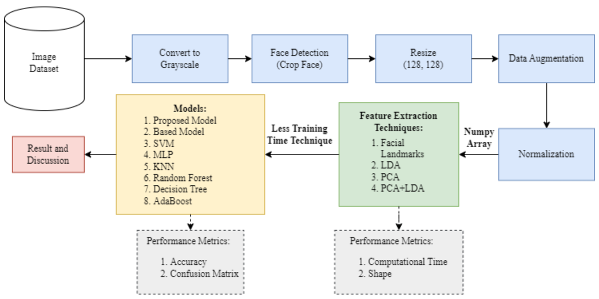
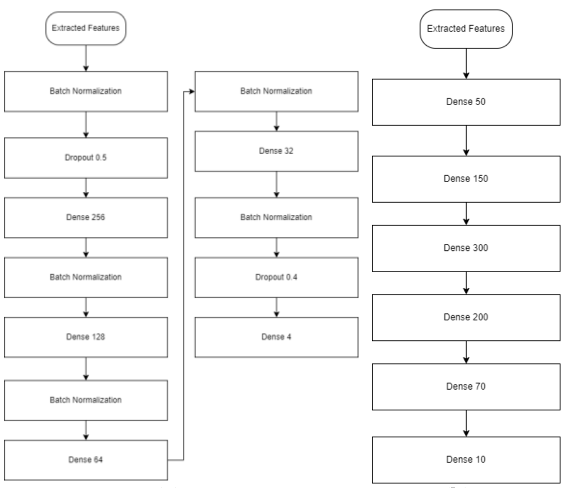


Fig. 2. Flow of the Experimental Setup



(a) (b)

Fig. 3. (a) Proposed Model Architecture, (b) Based Model Architecture

The outline of the experimental setup is shown in Fig. 2, The image is converted to grayscale and followed by face detection to detect and crop the face. Then it is resized into (128, 128) and undergoes data augmentation which consists of random rotation, random flip, random brightness and random blur. Next, the data will undergo normalization based on the feature extraction approaches and store in a NumPy array.

LDA, PCA, PCA+LDA uses Standard Scaler while facial landmarks localization uses histogram equalizer to normalize data. Then the extracted data is fitted into corresponding feature extraction approaches. In this case, LDA, PCA, PCA+LDA will use their built-in functions provided in sklearn while facial landmarks localization will use the function in dlib library to predict and extract the position of the landmarks using 68 landmarks predictor. The extracted data will be evaluated in terms of computational time and dimensionality shape.

After that, the feature extraction with the least computational time found in the previous stage will be implemented in the model classification.

Fig. 3(a) shows the proposed model layout architecture while Fig.3(b) shows the based model layout architecture. These 2 figures clearly depict the modification made from the referenced model from the proposed model. The proposed model adds 5 Batch Normalization layers and 2 Dropout layers with 0.5 and 0.4 at the beginning and in the end of the layout to prevent overfitting. The Dense layers' value is decremented by dividing by 2 to ensure the proposed model was stable. Both models used 'ReLu' activation function while the output dense layer 'softmax' activation function.

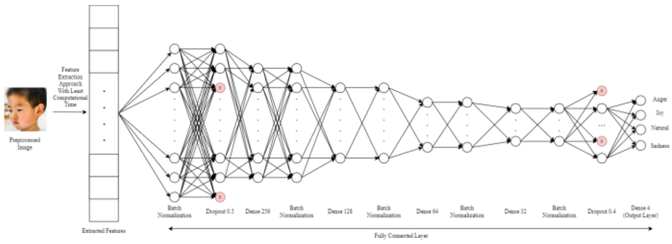


Fig. 4. Feature Extraction Approach Continue with Fully Connected Layer

Based on Fig.3(a), which describes the proposed model's layout architecture, Fig.4 shows more in details to explain the whole workflow of the proposed model. The input shape is the shape of the extracted feature shape followed by a fully connected layer, which consists of 5 Batch Normalization layers, 5 dense and 2 dropout layers. Convolutional and max pooling layers are excluded because the feature extraction has been carried out. The batch normalization layers are used to prevent overfitting, the same goes to the first dropout with 0.5 before the first dense layer and the last dropout with 0.4 before the last dense layer. A L2 regularizer with 0.0005 is added in the parameters of the first dense layer to increase the learning rate and prevent overfitting as well.

The features extracted are then fitted into the 7 classifiers. However, it cannot achieve to make every classifier to have the same parameters because of their architecture. Therefore, the classifiers' hyperparameters are either set to default or set to the popular-used parameters in facial emotion classification task. For the proposed model, based model and MLP, they used the same value of hyperparameters, such that learning rate = 0.0010, 50 epochs, 32 batch size and shuffle set to true. The callbacks used were early stopping with 10 patience and monitoring with validation accuracy. Another callback was used was ReduceLROnPlateau with 5 patience and monitoring with validation accuracy as well with 0.5 factor and 1e-6 minimum learning rate.

Then it comes to the evaluation based on the classification accuracy and confusion matrix. Finally, it will go to the result and discussions next.

V. Result And Discussions

This section analyzes and discusses the results obtained in the experiment.

TABLE I. Feature Extraction Result

Approach	Facial Landmarks Localization	LDA	PCA	PCA + LDA
Computational Time	0.8962	1.1306	0.2328	0.4371
Dimensionality Shape	(625, 268)	(416, 3)	(416, 64)	(416, 3)

Based on Table 1 which records the mean computational time and dimensionality shape, PCA consumed the least amount of computational time with 0.2328s and the dimensionality shape of (416, 64). Meanwhile, LDA took the longest computational time with 1.1306s and the dimensionality shape of (416, 3). PCA+LDA was the second approach that consumed least computational time with 0.4371s and the dimensionality shape of (416, 3). Meanwhile, facial landmarks localization took 0.8962s with (625, 268) dimensionality shape.

PCA was the least computational time as it involves matrix operations without iterative processes. Meanwhile, LDA required more computational time to compute a generalized eigenvalue. For the combination PCA and LDA, PCA reduced the dimensionality of the data first and passed to LDA to fit transform the data. The execution time was lesser than LDA because the shape was reduced by PCA, and LDA did not need to fit transform the whole bunch of data. The computational time taken was reasonable for facial landmarks localization as it implemented 68 landmarks predictors, which needed to scan through the image and predict and record the landmarks' positions. In addition, it also involved calculating the center deviance between the landmarks.

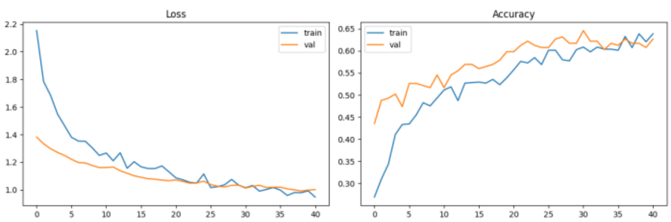


Fig. 5. Proposed Model's Loss and Accuracy Graphs

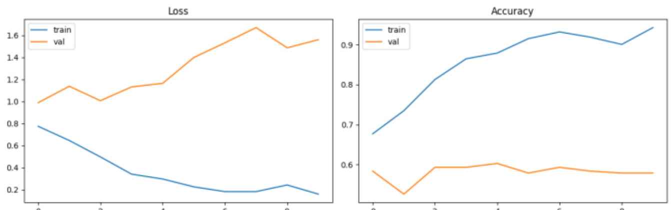


Fig. 6. Based Model's Loss and Accuracy Graphs

Fig. 5 and Fig. 6 show the loss and accuracy graphs for the proposed model and the referenced model respectively. The graphs in Fig. 6 show the based model was overfitting and stopped at epoch 10 because of Early Stopping. Meanwhile, the validation loss curve for the proposed model dropped from 1.4 to 1.0 smoothly while the accuracy loss curve dropped from 2.2 to 0.9 with a little fluctuation. This fluctuation was due to the added regularizer to prevent overfitting and improve generalization. In addition, the learning stopped at epoch 40 and the difference between both loss curves was not significant, which indicated that the proposed model was not overfitting. For the accuracy vs validation accuracy graph, the training accuracy increased from 0.25 to 0.65 while the validation accuracy increased from 0.43 to 0.62. The validation accuracy became consistent starting from 30. It was clearly shown that the proposed model's reliability was improved for the based model.

TABLE II. Model Classification Result

Classifier	**Accuracy**	**Precision**				**Recall**				**F1-Score**			
		A	***J***	***N***	***S***	***A***	***J***	***N***	***S***	***A***	***J***	***N***	***S***
Proposed model	64.30	0.52	0.78	0.54	0.50	0.16	0.89	0.44	0.62	0.25	0.83	0.48	0.56
Based Model	52.06	0.07	0.57	0.24	0.50	0.07	0.91	0.16	0.38	0.01	0.69	0.18	0.38
SVM	56.46	0.24	0.82	0.38	0.42	0.21	0.82	0.38	0.45	0.23	0.82	0.38	0.43
MLP	63.64	0.41	0.86	0.44	0.48	0.23	0.89	0.41	0.58	0.29	0.88	0.42	0.52
KNN	50.72	0.33	0.69	0.25	0.45	0.33	0.75	0.34	0.30	0.33	0.72	0.29	0.36
Random Forest	59.71	0.56	0.67	0.53	0.45	0.20	0.95	0.23	0.46	0.29	0.79	0.31	0.45
Decision Tree	47.18	0.19	0.70	0.33	0.37	0.18	0.67	0.38	0.37	0.18	0.68	0.35	0.37
AdaBoost	52.63	0.50	0.72	0.38	0.36	0.15	0.73	0.34	0.52	0.23	0.72	0.36	0.42

Where A = angry, J = joy, N = natural and S = sad

Table 2 records the average classification accuracy, precision, recall and f1-score. The proposed model had the best accuracy among the classifiers with 64.30%, which improved the accuracy of the based model (52.06%). Based on each emotion class's precision, recall and f1-score, it could be concluded that all classifiers learned well to predict joy emotion and did not learn well to predict the others. This was portrayed in the recall values of angry, natural and sad from all classifiers. This was due to the number of images in angry, natural and sad were less than that of joy.

VI. Conclusion And Future Works

Based on the result, it was concluded that PCA is the feature extraction approach that consumes the least computational time compared to facial landmarks localization, LDA and PCA+LDA. The proposed model, with 64.30% accuracy, performed better than the other classifiers, such as SVM, MLP, KNN, Random Forest, Decision Tree and AdaBoost. The relatively smaller value of recall in angry, sad and natural highlights that the classifiers do not learn well to classify these 3 emotions. In contrast, joy has higher recall values, indicating that it has greater number of joy images compared to the others in the dataset. Although classes like fear and disgust are dropped to prevent overfitting problem, biased classification inevitably occurs.

This research successfully found out that PCA is the feature extraction approach with the least computational time among the approaches. Furthermore, this work successfully proposed a way to recognize the autistic emotions such as angry, joy, natural and sad with higher classification accuracy compared to the classifiers. The reliability of the proposed model has been proven through the loss vs validation loss and accuracy vs validation accuracy graphs.

The future work is to collect more dataset of autistic facial emotions to increase the dataset size to improve the accuracy. Furthermore, finding an alternative way to increase the model classification accuracy using small dataset.

References

[1] H. Hodges, C. Fealko, and N. Soares, "Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation," *Translational Pediatrics*, vol. 9, no. S1, pp. S55–S65, Feb. 2020, doi: 10.21037/tp.2019.09.09. Available: https://tp.amegroups.com/article/viewFile/30253/pdf

[2] D. W. Black MD and J. E. G. M. D. , MPH JD, *DSM-5 Guidebook*, 5th ed., vol. 3. 2014.

[3] S. R. Sharma, X. Gonda, and F. I. Tarazi, "Autism Spectrum Disorder: Classification, diagnosis and therapy," *Pharmacology & Therapeutics*, vol. 190, pp. 91–104, Oct. 2018, doi: 10.1016/j.pharmthera.2018.05.007

[4] R. Brewer *et al.*, "Can Neurotypical Individuals Read Autistic Facial Expressions? Atypical Production of Emotional Facial Expressions in Autism Spectrum Disorders," *Autism Research*, vol. 9, no. 2, pp. 262–271, Feb. 2016, doi: 10.1002/aur.1508. Available: https://doi.org/10.1002/aur.1508

[5] E. García-Martín, C. F. Rodrigues, G. Riley, and H. Grahn, "Estimation of energy consumption in machine learning," *Journal of Parallel and Distributed Computing*, vol. 134, pp. 75–88, Dec. 2019, doi: 10.1016/j.jpdc.2019.07.007. Available: https://doi.org/10.1016/j.jpdc.2019.07.007

[6] R. K. Kana, *The Neuroscience of Autism*. Academic Press, 2022.

[7] M. J. Carter, "Diagnostic and Statistical Manual of Mental Disorders, 5th ed.," *Therapeutic Recreation Journal*, vol. 48, no. 3, Jan. 2014.

[8] P. Rani, Emotion Detection of Autistic Children Using Image Processing. 2019. doi: 10.1109/iciip47207.2019.8985706

[9] C.-C. Lee, E. Mower, C. Busso, S. Lee, and S. Narayanan, "Emotion recognition using a hierarchical binary decision tree approach," *Speech Communication*, vol. 53, no. 9–10, pp. 1162–1171, Nov. 2011, doi: 10.1016/j.specom.2011.06.004. Available: https://doi.org/10.1016/j.specom.2011.06.004

[10] Y. Wang, Y. Li, Y. Song, and X. Rong, "Facial expression recognition based on random forest and convolutional neural network," *Information*, vol. 10, no. 12, p. 375, Nov. 2019, doi: 10.3390/info10120375. Available: https://doi.org/10.3390/info10120375 H. Hodges, C. Fealko, and N. Soares, "Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation," *Translational Pediatrics*, vol. 9, no. S1, pp. S55–S65, Feb. 2020, doi: 10.21037/tp.2019.09.09. Available: https://tp.amegroups.com/article/viewFile/30253/pdf

[11] E. Owusu, Y. Zhan, and Q. R. Mao, "A neural-AdaBoost based facial expression recognition system," *Expert Systems With Applications*, vol. 41, no. 7, pp. 3383–3390, Jun. 2014, doi: 10.1016/j.eswa.2013.11.041. Available: https://doi.org/10.1016/j.eswa.2013.11.041

[12] H. I. Dino and M. B. Abdulrazzaq, "Comparison of four classification algorithms for facial expression recognition," *PolyTechnic/PolyTechnic (Erbil. Online)*, vol. 10, no. 1, pp. 74–80, Jun. 2020, doi: 10.25156/ptj.v10n1y2020.pp74-80. Available: https://doi.org/10.25156/ptj.v10n1y2020.pp74-80

[13] A. Mollahosseini, D. Chan, and M. H. Mahoor, *Going deeper in facial expression recognition using deep neural networks*. 2016. doi: 10.1109/wacv.2016.7477450. Available: https://doi.org/10.1109/wacv.2016.7477450

[14] J. O. Silva, I. L. Oliveira, Z. H. Topolniak, and A. G. Alvarez, *A CNN Approach Implemented to Emotional Facial Expression Recognition for the Prevention of Autistic Meltdowns*. 2021. doi: 10.1109/scla53004.2021.9540183. Available: https://doi.org/10.1109/scla53004.2021.9540183

[15] P. Marasamy and S. Sumathi, *Automatic recognition and analysis of human faces and facial expression by LDA using wavelet transform*. 2012. doi: 10.1109/iccci.2012.6158798. Available: https://doi.org/10.1109/iccci.2012.6158798

[16] S. Varma, M. Shinde, and S. S. Chavan, "Analysis of PCA and LDA features for facial expression recognition using SVM and HMM classifiers," in *Springer eBooks*, 2019, pp. 109–119. doi: 10.1007/978-3-030-16848-3_11. Available: https://doi.org/10.1007/978-3-030-16848-3_11

[17] L. Sun, J. Dai, and X. Shen, *Facial emotion recognition based on LDA and Facial Landmark Detection*. 2021. doi: 10.1109/icaie53562.2021.00020. Available: https://doi.org/10.1109/icaie53562.2021.00020

[18] F. M. Talaat, "Real-time facial emotion recognition system among children with autism based on deep learning and IoT," *Neural Computing and Applications*, Mar. 2023, doi: 10.1007/s00521-023-08372-9. Available: https://link.springer.com/content/pdf/10.1007/s00521-023-08372-9.pdf

[19] F. Mashuque Alamgir, S. Mostasim Hasnain Saif, S. Md. Hossain, A. Al Hadi, and Md. Shafiul Alam, "Facial Expression Database of Autism Spectrum Disorder Children," *European Chemical Bulletin*, Apr. 2023, doi: 10.48047/ecb. Available: https://doi.org/10.48047/ecb