Facial Expression Database of Autism Spectrum Disorder Children

Fakir Mashuque Alamgir [1*], Syed Mostasim Hasnain Saif [2], Saffat Md. Hossain [3], Abdullah Al Hadi [4], Dr. Md. Shafiul Alam [5]

[1]Ph.D. Student, Department of Electrical & Electronic Engineering,
University of Dhaka, Nilkhet Rd, Dhaka 1000, Bangladesh.
Assistant Professor, East West University, Dhaka, 1212, Bangladesh
[2] Department of Electrical & Electronic Engineering,
East West University, Dhaka, 1212, Bangladesh.
[3] Assistant Engineer, Dhaka Power Development Board.
[4] Research Assistant, East West University, Dhaka, 1212, Bangladesh.
[5] Professor, Department of Electrical and Electronic Engineering,
University of Dhaka, Bangladesh
Email: [1] fma@ewubd.edu

Abstract

The processing of face information relies on the quality of data resources and therefore the dataset is crucial for image processing. While considering behavioural investigations of facial expression recognition (FER) in autism spectrum disorders (ASD) provide conflicting findings due to its difficulty in processing real-world application domains. Also, the significant intra-class heterogeneity makes the FER a challenging task. Hence, this research introduces a novel facial expression database of ASD children. The database is gathered from ASD children aged 6 to 14 with various levels of disease severity, wherein 45% of the children have severe ASD, 35% have moderate, and 20% have low severity of ASD. A total of 81 Male and 32 female children participated in taking the images of facial expressions. Four expressions are considered when generating the dataset: Happy, Sorrow, Neutral, Angry or Disgusted. The generated dataset is validated by analyzing the facial expression recognition using the deep convolutional neural network (DeepCNN). The accuracy accomplished by the DeepCNN using the proposed ASD facial expression recognition database is 96.14%.

Keywords: ASD facial expression recognition, autism spectrum disorder children, landmark detection, deep convolutional neural network, ASD database, disease severity, facial expression.

1. Introduction

According to Kanner, one of the main symptoms of autism spectrum disorder (ASD) is impaired affective interaction [1]. For impaired affective interaction, it is crucial to recognize facial emotion [2]. A prior meta-analysis discovered that people with ASD have trouble recognizing facial emotions, and many studies have described abnormal processing of facial cues or problems with real-life emotional detection in people with ASD [3]. Functional neuroimaging research in healthy participants has extensively examined the brain processes of facial emotion recognition [4]. Earlier studies have shown that certain brain regions are organized hierarchically, with the visual cortex's activities corresponding to the processing of lower-level sensory information, such as the features of faces, and the superior temporal gyrus' or fusiform gyrus' activity patterns corresponding to the emotion category [5]. Additionally, clinical research has shown that ASD is associated with abnormal facial expression recognition network activity patterns.

The developmental mechanism of classifying visual facial expression information into emotional groups and its variations in ASD are mostly unclear [6]. Though the consistency of face feature patterns led to the theory that establishing emotional categories is a normal process. Still, the information processing by considering the variations in ASD in actual neural systems has rarely been examined [7]. To close the gap between the corresponding scientific discoveries and psychiatric symptoms, the area of computational psychiatry is developing to study information processing in the brain. A computational technique is utilized to comprehend face emotion detection and its impairment in ASD [8].

Facial landmark localization is useful for capturing cognitive facial expressions in computer vision. Identifying facial landmarks to understand the affective messages sent by facial emotions [9]. Facial landmarks are the placements of critical facial features crucial for recognizing the expression. Using face landmarks from image sequences, geometric facial features are extracted to help identify facial expressions [10, 11]. Effective classification of the six types of facial expression states requires a computationally efficient facial expression recognition system that evaluates the active patches and identifies the prominent spots on the face [12]. A 3-D statistical facial feature model also captures local differences in texture and geometry around each landmark and global variations in configurational links between landmarks [13].

Facial landmark localization is an efficient method for cognitive facial expression in computer vision [14]. Facial landmark localization localizes the locations of the crucial parts of a face and are important for the perception of the affective conveyance of facial expressions [15]. The geometric facial feature extraction by tracking facial landmarks of image sequences is also essential for facial expression recognition. A computationally efficient facial expression recognition system analyses the active patches and determines the salient areas on the face for accurate classification of the six classes of facial expression states [16]. Moreover, a 3-D statistical facial feature model learns the global variations in configurational relationships between landmarks and the local variations of texture and geometry around each landmark [17].

Due to its ability to anticipate outcomes accurately, deep learning is widely used in various application fields like visual aid, healthcare, and entertainment [18]. A deep learning model's generalization capability helps to recognize and accurately forecast the pattern of previously unknown data or new data derived from the same distribution as the training data [19]. Here, the generalization describes a model that acquires new data after being trained on a training dataset and makes accurate predictions using the unknown input data [20]. Hence, the dataset is crucial for the deep learning-based application domains. At the same time, considering facial expression recognition for ASD persons, the availability of a dataset is insufficient. Hence, a novel dataset concerning ASD children is presented in this research. The major contributions of the research are:

- ***Image Acquisition***: The facial expressions of ASD children under the 6-14 age group were captured using the Canon M50 camera with a 24.1 megapixel APS-C CMOS sensor and an EF-M mount.
- ***Image Selection***: Human raters were used to select images with good intensity and purity for the facial expressions categorization.
- ***Image Categorization based on expression***: Using the facial landmark technique based on the Dlib toolkit, the database organization based on four classes like Happy, Sorrow, Neutral, Angry or Disgust is devised.
- ***Validation***: The validation of a database is analyzed by evaluating the proposed ASD-FER database using the DeepCNN based on various assessment measures like accuracy, sensitivity, precision and recall.

The organization of the research are: Section 2 details the related FER dataset along with the challenges. Section 3 details the dataset generation and the validation of the database. Finally, Section 4 concludes the research.

2. FER Database

The conventional FER database is analyzed in this section. The use of overly simplistic stimuli, such as 100% full-blown or expression can induce ceiling effects and is considered one of the causes for the extremely inconsistent results on the difficulty with FER in ASD children [21]. Individuals with ASD provide excessive emotions in facial expressions that are more authentic and indicative of actual feelings [22]. Due to the fact that subtle facial expressions provide fewer emotional indicators to the spectator, which could exhibit greater difficulty in recognizing emotions when the expressions are low intensity. Also, low-intensity emotions like negative expressions resulted in FER deficits. In everyday life, it is typically involves noticing low-intensity or subtle facial expressions. Children with ASD frequently perform well in FER in a lab context, but in the real world, it is likely to struggle for recognizing the nuanced facial expressions [23]. Thus, there is probable that the incorrect processing of facial emotions employing high intensity of stimuli accounts for certain studies' findings that FER is not impaired in people with ASD [24].

A. Problem Statement

Social interaction through non-verbal communication is the key communication technique devised through facial expressions [3]. From facial expressions, one can understand the other and decode the information [9]. Thus, facial expression allows for identifying emotional states that help foster social interaction, at least partially among individuals who cannot communicate verbally [21]. Also, the development of empathetic feelings and the Theory of Mind based on mental state features state that facial expressions promote regulation and adaptation of interaction through the behaviour [25]. Several types of research have demonstrated that people with ASD have a lifetime of difficulty reading and comprehending others' facial expressions [26]. Although the problems with emotion processing appear to be widespread, numerous studies have shown that they are particularly likely to impact the recognition of negative facial expressions such as surprise, sadness, anger, fear, and contempt [20]. Hence, this research introduces the facial expression recognition of ASD children aged 6-14.

3. ASD Facial Expression Database Generation

The system set up is made initially to capture images of ASD children with various facial expressions. Then, the images of facial expressions for the children are captured. From the gathered images, the human raters choose the image with better intensity and purity to select the best quality images. Using the facial landmark technique based on the Dlib toolkit, the database organization based on four classes like Happy, Sorrow, Neutral, Angry or Disgust is formulated. Finally, the validation of an introduced database is analyzed using the DeepCNN for analyzing the performance. The database generation process is depicted in Figure 1.

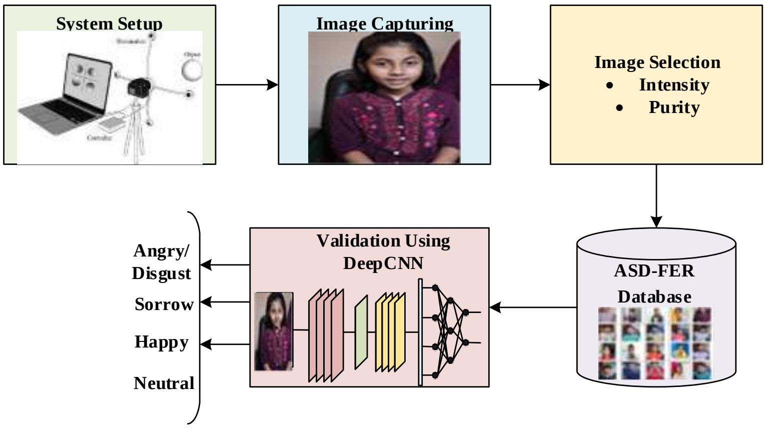

Figure 1: Database Generation Process

A. System Setup

The facial expression images are captured using the Canon M50 camera with a 24.1 megapixel APS-C CMOS sensor and an EF-M mount. The lens utilized in the camera is Canon EF-M Mount 15-45mm f/3.5-6.3 zoom lens, a kit lens often bundled with the M50. The developed system captures the facial expressions of Autism spectrum disorder (ASD) children. The children are asked to sit in a comfortable chair and capture the image during the image acquisition process. Also, the ceiling lights were kept OFF, and no flashes in the camera were utilized. For smooth and soft illumination, three spotlights based on professional studio photography are utilized, wherein the softening of the light is acquired by utilizing white photography umbrellas. Here, two lights are kept at 45 degrees lit from two sides of the child and one above the height of 50cm above the child's midline.

B. Image Capturing

The ASD children, aged 6-14, participated in capturing images for facial expression recognition. Totally, 81 Male and 32 female children were involved in recording facial expressions. Four expressions, Happy, Sorrow, Neutral, Angry or Disgusted, are taken to generate the dataset. In the proposed generation of children's ASD facial expression datasets, all images were captured in portrait orientation, which is most suitable for capturing the children's facial expressions. The size of the images varied from 5-10 megabytes, which is reasonable for a 24.1-megapixel camera. While taking the images, cosmetics such as lipstick, blush, eye shadows and others related to varying the skin texture are not allowed. The resolution of the images is $4000 * 6000$ and saved for further processing.

C. Image Selection

The best image based on the ASD-based facial expression concerning the positive, negative or neutral expression is chosen from the image captured in the previous phase. Initially, the out-of-focus image, blurred and ambiguous facial expression images are removed due to the poor quality of images. Then, the image selection based on intensity and purity is devised in this stage, which is employed manually by the human raters.

Step 1: The raters refer to the manual developed by Ekman & Friesen [27] to select the best images for generating the dataset.

Step 2: The raters must define the purity and its intensity (high, low and medium) to select better-quality images. Two raters are assigned for each image, and the image with a better rating by both raters is chosen for further processing.

Step 3: The repeated similar images are removed finally in the image selection process.

D. Database Organization

The database organization is performed based on four classes: Happy, Sorrow, Neutral, Angry or Disgusted. Here, the organization of the dataset is employed using the facial landmark detection approach. The images selected by the raters are utilized to organize facial expressions. The facial variations are differentiated based on the centre of a mouth, contour points of a mouth, nostrils, tips of a nose, pupils, eye contour points, and eyebrows. In ASD facial expression recognition, the Dlib tool kit [28] represents the facial landmark localization to categorize the expressions more accurately. The Dlib toolkit is the C++ based algorithm widely utilized to solve real-world issues. 68 landmarks are used to detect facial expressions based on four classes. The landmark localization using the Dlib toolkit is depicted in Figure 2.

Figure 2: Landmark localization by Dlib toolkit

The Samples of the landmark localization for four various classes using the Dlib toolkit are depicted in Figure 3.

Expression	Image Samples	Landmark Localization

Happy		
Sorrow		
Neutral		
Disgust		

Figure 3: Example of landmark localization for four various classes

1) Description of Dataset

The facial expression recognition dataset for ASD children aged 6-14 is categorized into four classes: Happy, Sorrow, Neutral, and Angry or Disgusted. 113 images are gathered for generating the dataset with 1 or 2 images for each subject. The autism level of the children is 20%, 35% and 45% based on the low, moderate and severe categories. The detailed dataset description is presented in Table 1.

Table 1: Dataset Description

Expression	Male	Female	Total
Happy	25	10	35
Sorrow	16	5	21
Neutral	27	13	40
Angry/Disgust	13	4	17
Total	81	32	113

2) Samples of Dataset Organization

The image samples after organizing the dataset with four classes: angry/disgusted, happy, sad, and neutral, are illustrated in Figure 4.

Figure 4: Sample images from the organized dataset

E. Validation

The validation of a proposed ASD facial expression recognition dataset is employed by processing the facial expression recognition using a deep convolutional neural network (DeepCNN).

1) Architecture of DeepCNN:

Deep learning models consist of multiple layers to learn the complex attributes of the input data to make the generalization more accurate. Besides, the deep learning models offer promising solutions for application domains like natural language processing, signal processing, topic classification, speech recognition, etc. Several deep learning methods, such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Deep Belief Networks (DBN) and several other methods, have been developed to process various applications. Deep CNN (DCNN) is widely used in various computer vision applications due to better performance through the data learning process. In addition, DeepCNN's automatic attribute extraction criteria eliminate the need for manual feature extraction. Therefore, the DeepCNN is used in the proposed facial expression validation. The DeepCNN consists of five fully connected layers, a max-pooling layer, two 1D convolution layers, and the flattening layer for detecting the facial expression of ASD children, as shown in Figure 5.

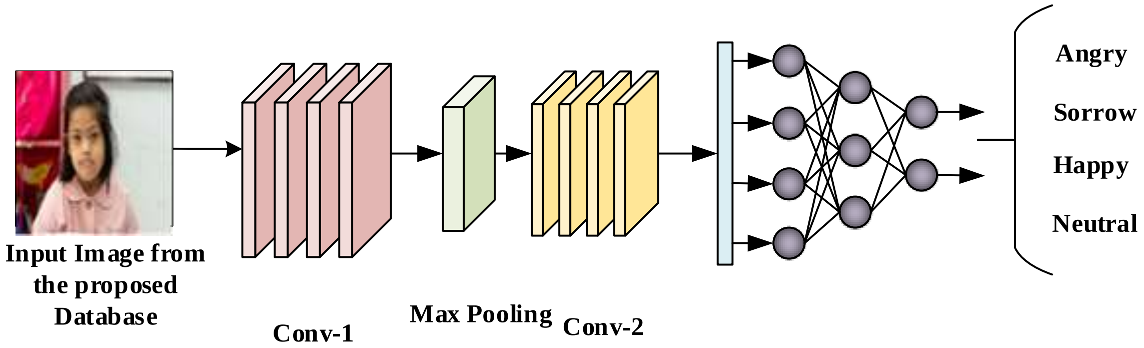

Figure 5: Architecture of DeepCNN

Conv Layer: Several kernels constitute the convolutional layers of DeepCNN that perform the linear operation to create a mapping of an input image. Here, the proposed ASD-FER database serves as the source for the input used for FER. Due to kernel size change, information loss may occur. However, the issue concerning information loss could be avoided by employing the information padding technique. The convolutional layer's output is outlined as follows:

$$FER_{conv} = \sum X^{w} * Y^{w} + Q^{w} \tag{1}$$

where, the bias is indicated as Q^w, in which w refers to the w^{th} output map, the weight is indicated as Y^w, the input feature is indicated as X^w, and the output of a conv layer is indicated as FER_{conv}.

Max-Pooling Layer: The fine and strong characteristics are retrieved from the convolved feature map using the max pooling operation. To reduce computing costs, the max-pooling layer is employed in the DeepCNN. The most activated features are extracted using the max-pooling layer of the DeepCNN.

Flatten Layer: The flatten layer transforms the characteristics into a single dimension.

Fully Connected Layer: The DeepCNN's fully connected layer includes a softmax activation function to categorize various facial expressions. The formula corresponding to the classification is written as follows,

$$FER_{out} = \frac{e^{z_m}}{\sum_{n=1}^{i} e^{z_n}}, \qquad m = 1,2,3,4 \tag{2}$$

where, the softmax function is indicated as FER_{out}, the element corresponding to an input attribute is indicated as z_m , and i refers to the number of classes in the output. Here, the number of classes is four; hence the value of m is assigned with the variable (1, 2, 3, 4).

2) Assessment using the proposed database

The Assessment of a proposed ASD-FER database using the DeepCNN based on various assessment measures like accuracy, precision, recall and F-Score is depicted in Figure 6, and its detailed Assessment is depicted in Table 2. Here, the maximal accuracy estimated by the DeepCNN is 96.14%, the maximal precision is 94.03%, the maximal recall is 91.43%, and the maximal F-Measure is 93.39%. Here, the analysis is devised by using 80% data for training the DeepCNN and 20% of the data for testing.

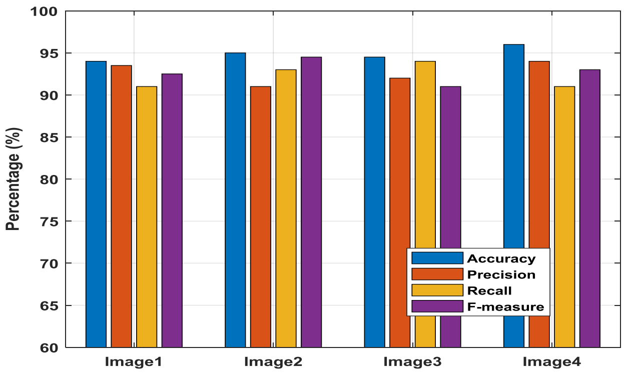

Figure 6: Assessment of the proposed ASD-FER database using the DeepCNN

Table 2: Assessment of the proposed ASD-FER database

Methods/ Image Samples	Image1	Image2	Image3	Image4
Accuracy (%)	94.32	95.09	94.5	96.14
Precision (%)	93.5	91.42	92.57	94.03
Recall (%)	91.01	93.78	94.91	91.43
F-measure (%)	92.5	94.5	91.26	93.39

F. Discussion

In recent years, computer vision and deep learning methods have been widely utilized for image processing application domains. Thus, noticeable progress has been made in classifying and recognizing FER deep learning methods. The FER devised with seven facial expressions like anger, disgust, fear, sad, contempt, surprise, and happiness using the FER2013 dataset acquired 71.16% test accuracy by the "Challenges in Representation Learning: Facial Expression Recognition Challenge". Moreover, Convolutional Neural

Networks (CNNs) are widely used for image classification and recognition tasks. Thus, the ASD-FER is tested using the DeepCNN model to depict the test accuracy of the proposed ASD-FER data. The enhanced accuracy of a model depicts the better organization of a dataset that enhances the generalization capability of the classifier. Children with ASD are the intended participant for the dataset generation to recognize and categorize facial expressions of them in real-time processing. At the same time, they practice social skills during routine social interaction.

G. Limitations

Some limitations of the gathered ASD-FER dataset are:

- The proposed ASD-FER dataset generation is devised by gathering images from relatively few participants. Thus, FER analysis, which requires a large amount of data, is not possible with the proposed database. Also, there is a need for detailed database analysis regarding various facial expressions concerning ASD children, like fear, surprise and so on.
- The facial expressions of ASD children vary based on the type of disorder, which can be measured using standardized scales. Psychological disorders like depression, anxiety, and mood disorder need to be included in the database to acquire full medical data on ASD children. Thus, the absence of disorder inclusion is the limitation of a technique.

4. Conclusion

This research introduces a novel dataset for the FER of ASD children. Images from ASD children aged 6-14 are gathered initially and then rated using human raters based on purity and intensity to select the best-quality images. Then, using the Dlib toolkit, the facial expressions are categorized into four expressions: Happy, Sorrow, Neutral and Angry/Disgusted. Finally, the validation is devised using DeepCNN, and the outcome is evaluated based on accuracy, precision, recall and F-Measure. The analysis using the proposed ASD-FER dataset accomplishes the maximal accuracy, precision, recall and F-Measure of 96.14%, 94.03%, 91.43%, and 93.39%, respectively. Normally, deep learning methods improve accuracy by learning enormous data. The higher amount of data learning elevates the generalization capability, which in turn enhances the accuracy while testing the unknown data. Due to the usage of minimal data for analysis, the accuracy of DeepCNN in FER is limited. Hence, in the future large amount of data will be gathered to enrich the dataset.

References

[1] Y. Takahashi, S. Murata, H. Idei, H. Tomita and Y. Yamashita, “Neural network modeling of altered facial expression recognition in autism spectrum disorders based on predictive processing framework”. Scientific reports, vol. 11, no. 1, pp. 14684, 2021.

[2] M. Ventura, F. Innamorato, A. Palmisano, G. Cicinelli, E. Nobile, V. Manippa, R. Keller and D. Rivolta, “Investigating the impact of disposable surgical face- masks on face identity and emotion recognition in adults with autism spectrum disorder”. Autism Research 2023.

[3] A.A. Alhussan, F.M. Talaat, E.S.M. El-kenawy, A.A. Abdelhamid, A. Ibrahim, D.S. Khafaga and M. Alnaggar, "Facial Expression Recognition Model Depending on Optimized Support Vector Machine."
[4] U. Laraib, A. Shaukat, R.A. Khan, Z. Mustansar, M.U. Akram and U. Asgher, "Recognition of Children's Facial Expressions Using Deep Learned Features". Electronics, vol. 12, no. 11, pp. 2416, 2023.
[5] S. Li and W. Deng, "Deep facial expression recognition: A survey". IEEE transactions on affective computing, vol. 13, no. 3, pp.1195-1215, 2020.
[6] U.M. Pandyan, M.M.R. Sindha, P. Kannapiran, S. Marimuthu and V. Anbunathan, "Application of Machine and Deep Learning Techniques to Facial Emotion Recognition in Infants" 2023.
[7] D. Zeng, Z. Lin, X. Yan, Y. Liu, F. Wang and B. Tang, "Face2exp: Combating data biases for facial expression recognition". In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 20291-20300, 2022.
[8] A.C. Dantas and M.Z. do Nascimento, "Recognition of emotions for people with autism: An approach to improve skills". International Journal of Computer Games Technology, vol. 2022, pp.1-21, 2022.
[9] C. Zheng, M. Mendieta and C. Chen, "POSTER: A Pyramid Cross-Fusion Transformer Network for Facial Expression Recognition" 2022. arXiv preprint arXiv:2204.04083.
[10] M.A. Haghpanah, E. Saeedizade, M.T. Masouleh and A. Kalhor, "Real-time facial expression recognition using facial landmarks and neural networks". In 2022 International Conference on Machine Vision and Image Processing (MVIP) IEEE pp. 1-7, 2022, February.
[11] M. Wafi, F.A. Bachtiar and F. Utaminingrum, "Feature extraction comparison for facial expression recognition using adaptive extreme learning machine". International Journal of Electrical & Computer Engineering vol. 13, no. 1, pp. 1113-1122, 2023.
[12] Y. Yaddaden, "An efficient facial expression recognition system with appearance-based fused descriptors". Intelligent Systems with Applications, vol. 17, pp.200166, 2023.
[13] P. Choirina, U.D. Rosiani, I.M. Fitriani and R. Baqi, "Facial Micro Expression Recognition for Feature Point Tracking using Apex Frames on CASME II Database". Sinkron: jurnal dan penelitian teknik informatika, vol. 8, no. 1, pp. 81-89, 2023.
[14] D. Kollias, "Multi-Label Compound Expression Recognition: C-EXPR Database & Network". In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 5589-5598, 2023.
[15] H. Laux, A. Hallawa, J.C.S. Assis, A. Schmeink, L. Martin and A. Peine, "Two-stage visual speech recognition for intensive care patients". Scientific Reports, vol. 13, no. 1, pp. 928, 2023.
[16] M. Hao, F. Yuan, J. Li and Y. Sun, "Facial expression recognition based on regional adaptive correlation". IET Computer Vision 2023.
[17] J. Wei, H. Chen, G. Lu, J. Yan, Y. Xie and G. Zhao, "Prior information based Decomposition and Reconstruction Learning for Micro-Expression Recognition" 2023. arXiv preprint arXiv:2303.01776.

[18] P. Washington and D.P. Wall, "A Review of and Roadmap for Data Science and Machine Learning for the Neuropsychiatric Phenotype of Autism". Annual Review of Biomedical Data Science, vol. 6, 2023.

[19] A. Rashid and S. Shaker, "Autism spectrum Disorder detection Using Face Features based on Deep Neural network". Wasit Journal of Computer and Mathematics Sciences, vol. 2, no. 1, 2023.

[20] F.M. Talaat, Z.H. Ali, R.R. Mostafa and N. El-Rashidy, "Real Time Facial Emotion Recognition model Based on kernel Autoencoder and Convolutional Neural Network for Autism Childrens" 2023.

[21] H.A. Valera and M. Luštrek, "Unsupervised Domain Adaptation for Facial Emotion Recognition in Autistic Children". In Workshops at 18th International Conference on Intelligent Environments (IE2022) IOS Press vol. 31, pp. 66, 2022, July.

[22] A. Landowska, A. Karpus, T. Zawadzka, B. Robins, D.E. Barkana, H. Kose, T. Zorcec and N. Cummins, "Automatic emotion recognition in children with autism: a systematic literature review". Sensors, vol. 22, no. 4, pp.1649, 2022.

[23] H. Meyer-Lindenberg, C. Moessnang, B. Oakley, J. Ahmad, L. Mason, E.J. Jones, H.L. Hayward, J. Cooke, D. Crawley, R. Holt and J. Tillmann, "Facial expression recognition is linked to clinical and neurofunctional differences in autism". Molecular Autism, vol. 13, no. 1, pp.43, 2022.

[24] S. Jelili, S. Halayem, A. Taamallah, S. Ennaifer, O. Rajhi, M. Moussa, M. Ghazzei, A. Nabli, S. Ouanes, Z. Abbes and M. Hajri, "Impaired recognition of static and dynamic facial emotions in children with autism spectrum disorder using stimuli of varying intensities, different genders, and age ranges faces". Frontiers in Psychiatry, pp.1432, 2021.

[25] C. Darwin, "The Expression of the Emotions in Man and Animals". John Murray, London 1872. http://dx.doi.org/10.1037/10001-000

[26] S. Uono, W. Sato, T. Kochiyama, S. Yoshimura, R. Sawada, Y. Kubota, M. Sakihama and M. Toichi, "The structural neural correlates of atypical facial expression recognition in autism spectrum disorder". Brain Imaging and Behavior, vol. 16, no. 3, pp.1428-1440, 2022.

[27] E.A. Clark, J.N. Kessinger, S.E. Duncan, M.A. Bell, J. Lahne, D.L. Gallagher and S.F. O'Keefe, "The facial action coding system for characterization of human affective response to consumer product-based stimuli: a systematic review". Frontiers in psychology, vol. 11, pp. 920, 2020.

[28] F. You, Y. Gong, H. Tu, J. Liang and H. Wang, "A fatigue driving detection algorithm based on facial motion information entropy. Journal of advanced transportation", vol. 2020, pp.1-17, 2020.