IDEmotion-GAN: Synthetic Data Augmentation for Facial Emotion Recognition in Intellectually Disabled Children

Surendra Ramteke [1], Dr. Sunil Kumar [2]
[1] Post Doctoral Researcher; [2] Associate Professor and Adjunct Research Faculty
Lincoln University College, Malaysia
Email ID : pdf.surendra@lincoln.edu.my , pdfsv.sunilkumar@lincoln.edu.my

Abstract: Facial Emotion Recognition (FER) has emerged as a crucial element in the fields of assistive technologies, healthcare systems, and intelligent educational platforms. Current FER models, however, are largely based on data from NONID adults, and are ineffective for children with ID, because of their limited sample sizes, atypical emotional expressions and high class imbalance. In this study, the authors introduce their data augmentation methods, namely data augmentation framework IDEmotion-GAN, which generates realistic synthetic facial images to enhance emotion recognition performance on intellectually disabled children. The framework combines three specialized datasets derived from autistic and intellectually disabled children and uses a Conditional Generative Adversarial Network (cGAN) to provide more examples of under-represented emotion categories. The augmentation pipeline consists of multiple steps, including rotation, reflection, translation, cropping, and GAN-based synthesis, to create a more diverse dataset. The experimental analysis shows that the proposed augmentation process is very effective in boosting the class balance and recognition accuracy than the traditional training methods. Emotion-specific characteristics in the synthetic samples are preserved while also increasing the intra-class variability for more robust training of the models. The proposed framework can be used in future to solve data scarcity problems in special needs emotion recognition systems, and can be useful for developing emotion-aware learning environments, emotion-aware therapeutic monitoring, and inclusive healthcare environments.

Keywords: Facial Emotion Recognition; Intellectual Disability; Data Augmentation; Conditional GAN; Synthetic Emotion Dataset

Introduction

Emotion recognition is a key component in human communication, social interaction and development of cognition. One of the most informative channels by which one expresses and comprehends emotions is facial expressions. Therefore, Facial Emotion Recognition (FER) systems have been of great interest in the last few years because of their use in the medical field, the education sector, Intelligent tutoring systems, human computer interaction, and assistive technologies [1].

Although significant strides have been made in the field of FER research, most of the current FER systems have been created from data gathered from neurotypical subjects. Many widely used databases like FER2013, AffectNet, RAF-DB and CK+ are generally composed of photos taken in a controlled environment and frequently depict adults with typical emotions [2]. Models that are trained from such datasets tend to perform well on like datasets, but not so well on neurodiverse datasets.

Intellectually disabled children may exhibit emotions in a different manner than non-disabled children. They might have subtle, inconsistent, delayed, and/or influenced expressions, depending on developmental disorders that may exist concurrently with their own, such as Autism Spectrum Disorder (ASD) and Attention Deficit Hyperactivity Disorder (ADHD) [3]. These special features pose great difficulty for conventional FER models which were designed to read standard facial patterns. Thus, emotion recognition systems built on the normal population may have poor performance in detecting emotions in intellectually disabled individuals.

One difficulty is the lack of publicly available databases of facial expressions of intellectually disabled children. Ethical, privacy, availability and expert supervision for data collection in this population present significant challenges. Typically, existing datasets are therefore small and imbalanced, and not suitable for the effective training of deep learning models [4]. When the training set is small, there is insufficient data diversity to learn well, resulting in overfitting, less generalization, and worse performance in the real world if there are emotional differences.

However, with recent developments in Generative Artificial Intelligence (GenAI), new avenues for tackling data scarcity issues have emerged. Generative Adversarial Networks (GANs) have shown incredible power to create realistic synthetic images that maintain salient visual features of the original image distributions [5]. Samples produced by GANs have been used to augment visual datasets for computer vision applications with great success to diversify the data, increase class balance and boost the robustness of the model. However, there is limited research on emotion recognition by augmentation using GAN in intellectually disabled children.

To overcome this research gap, the present study proposes a synthetic data augmentation framework specifically for facial emotion recognition in intellectually disabled children, which is called IDEmotion-GAN. The approach proposed combines three distinct sets of specialized data that are related to populations of neurodiversity and uses a Conditional Generative Adversarial Network (cGAN) to create more samples from under-represented categories of emotions. The framework not only extends the dataset with traditional augmentation methods but also incorporates synthetic image generation to further enrich the dataset, thereby enhancing model learning and the emotion recognition accuracy.

In contrast to common augmentation techniques, which simply change the appearance of an image by rotating it, moving it around, and flipping it upside down, the augmentation performed by GAN can produce completely new images while maintaining the emotion-specific facial features. This is very helpful when the number of actual samples of the minority emotion categories is small. This enriched and well-balanced data will help to better represent emotional expressions seen in intellectually disabled children.

The main goal of this study is to examine if synthetic data generation can alleviate data imbalance and enhance FER performance in the context of neurodiverse groups. The study also assesses the recognition accuracy achieved with various augmentation methods and analyzes their effects against various emotion categories.

This paper is organized as follows: Section 2 summarizes the latest research works of both emotion recognition in the face and GAN-based augmentation methods. The main results of this study are summarised in Section 3. The methodology, experimental analysis and results are described in subsequent sections.

Related work

In recent years, the issue of developing FER systems for special-needs populations has been gaining increasing interest. It has become clear that traditional emotion recognition models tend to be insufficient for people with atypical emotion behavior patterns.

Tanabe et al. [6] proposed physiological signals and eye-gaze based emotion recognition framework for children with profound intellectual and multiple disabilities. They used a heart rate variability data along with a random forest (RF) classifier and obtained a recognition rate of around 70%. The study illustrated the feasibility of multi-modal sensing but limited the study to binary classification of emotions and small number of participants.

Tamas et al. [7] examined emotion recognition abilities of people with autism spectrum disorders and intellectual disabilities. They found that there were significant challenges when it came to recognizing negative emotions like fear and anger, and that there is a need for specific frameworks of emotion recognition systems that are unique to neurodiverse populations.

Talaat [8] proposed a deep learning based FER system for autistic children with optimized Convolutional Neural Networks (CNN) with Internet of Things (IoT) infrastructure. The proposed framework achieved extremely high emotional classification accuracy for 6 emotional categories. The study was mostly based on facial images without considering data scarcity and class imbalance problems.

Operto et al. [9] studied the children and adolescent specific learning disorders' skills in facial emotion recognition. Their work showed that emotion recognition deficits were linked to cognitive/ executive functioning deficits, and not to overall cognitive ability. Although it is clinically relevant, the study didn't include classification mechanisms based on artificial intelligence.

Gaya-Morey et al. [10] studied the use of deep learning models for facial expressions recognition in people with intellectual disabilities. They concluded that the patterns of facial expression were significantly different in intellectually disabled populations as compared to neurotypical populations. The authors highlighted the need to develop models that are specific to the population of interest and are able to flex with these expression parameters.

Synthetic data generation has been shown to be effective solution to class imbalance issues in recent studies. Roy et al. [11] added to FER datasets using a diffusion-based image synthesis method, and showed significant results on FER2013 and RAF-DB benchmarks. They showed that synthetic augmentation was crucial in diversifying the dataset and increasing the accuracy of classification.

Likewise, the medical imaging, biometric analysis and affect computing fields have seen some use of GAN-based augmentation be reported as successful [12]-[14]. GANs allow for the creation of synthetic samples for training the models, which helps to learn richer representations and mitigate overfitting due to limited training data.There is ample evidence in the existing literature of the need for specialized FER systems for neurodiverse populations. However, not much consideration has been given to addressing the core issue of data scarcity. In the present effort, this is achieved by introducing a specific emotion recognition synthetic augmentation framework developed for an Intellectually Disabled Child (IDC).

Table 1. Comparison of Recent Emotion Recognition Studies

Study	Population	Dataset Type	Augmentation	Reported Accuracy
Tanabe et al. [6]	Profound ID Children	Physiological Signals	No	70.4%
Talaat [8]	Autistic Children	Facial Images	Limited	99.9%
Gaya-Morey et al. [10]	Intellectual Disability	Facial Images	No	Not Reported
Roy et al. [11]	General Population	FER2013, RAF-DB	Diffusion Models	96.4%
Proposed IDEmotion-GAN	Intellectually Disabled Children	Multi-source FER Dataset	cGAN + Traditional Augmentation	Improved Dataset Balance

Key Contribution

The key findings of this study are summarized below:

i. Multiple different specific datasets, each with their own facial expression data for autistic children and intellectually disabled children, are combined to form a unified facial emotion dataset.Multiple specific datasets, each comprising their own facial expression data for autistic and intellectually disabled children respectively, are formed to create a unified facial emotion dataset.

ii. To tackle the issue of class imbalance and insufficient data, a Conditional GAN-based synthetic image generation framework is developed.

iii. To improve the diversity of the dataset, samples are systematically mixed with traditional augmentation methods such as rotation, translation, cropping and reflection, in addition to the GAN-generated samples.

iv. The effectiveness of synthetic augmentation is evaluated and analyzed to enhance emotion recognition performance in various emotion categories.

v. The proposed framework provides a basis for creating more inclusive and extensive FER systems for neurodiversity.

Method, Experiments and Results

Dataset Description

One of the challenges of deep learning-based Facial Emotion Recognition (FER) systems is that the quality, diversity and balance of the training data plays a crucial role in the performance of the system. In this study, a combined emotion dataset was created by merging three specialized datasets on facial expression of children with autism and intellectually disabled children. The data sets were combined to reflect a broad spectrum of emotional responses seen in the neurodivergent community.

FERAC (Facial Emotion Recognition of Autistic Children) is the first one, which contains facial images that are taken under clinical supervision from children diagnosed with Autistic Spectrum Disorders. There are four main categories of emotion in the dataset: Natural, Fear, Joy and Anger. The second is Autistic Children Emotions that consist of six classes of emotions: Surprise, Delight, Sadness, Fear, Joy, Anger. The third data was obtained from the Utkarsha Matimand Vidyalaya which contains in it the natural expressions of the emotions of the intellectually disabled child in an educational environment.

These datasets merged created a comprehensive dataset of seven categories of emotion:Natural,Fear,Joy,Anger,Surprise,Sadness,Delight.Although the datasets were integrated, there was still a high degree of class imbalance. Whereas the numbers of samples representing certain emotions (Joy and Natural) were significantly higher than the others (Fear, Surprise, Sadness). This can cause deep learning models to be biased towards majority classes and result in lower emotional recognition rates for minority emotional categories. In order to overcome this limitation, an extensive framework for augmentation was developed, consisting of both conventional transformations and synthetic image generation.

Data Preprocessing

All the face images were pre-processed in the same manner before augmentation and model training.

i. Face Detection and Cropping

The facial regions were extracted by using Viola–Jones' face detection framework. The cropped and isolated face region was considered as a face image to remove irrelevant visual information from the background.

ii. Image Normalization

All the facial images retrieved were resized to a fixed size of 224 × 224 pixels. The intensity values of the pixels were normalized to guarantee steady training and convergence of deep learning models.

iii. Emotion Annotation Verification

Emotion labels were checked from the original datasets, and a common labeling format was used. This process ensured uniformity and consistency between all the datasets, and minimized discrepancies in the annotations.

The preprocessing stage generated a preprocessed and standardized dataset for augmentation and further analysis.

Conventional Data Augmentation

Multiple augmentation techniques were used for the traditional ones to increase the diversity of the data sets and strengthen the model robustness**.**

i. Random Rotation

To mimic natural head movements that are often seen in children during emotional interactions, images were rotated between ±15 degrees.

ii. Horizontal Reflection

Mirror images were created by performing horizontal flips on the image, with emotional semantics retained. This increased the number of training samples by a factor of 2 and enhanced the resilience to facial orientation.

iii. Translation

Images were also moved horizontally and vertically by small distances to simulate off-centre faces that are often seen in real world scenarios.

iv. Random Cropping

The partial face crops were created to simulate occlusion situations and motivate the model to learn discriminative local features.

v. Brightness Adjustment

Brightness and contrast variations were added to simulate different illumination conditions in real world situations.

This is due to the various changes that have been made which caused a high intra-class variance but maintained emotion-specific features.

A conditional GAN-based synthetic data generation method

Traditional augmentation will add more variability to the set of data, but it cannot create new patterns of emotion. Thus, a Conditional Generative Adversarial Network (cGAN) was used to generate realistic emotion samples of the face.

There are two competing neural networks in the cGAN framework:

Generator Network

The generator takes an input of a random noise vector, a label of emotion and generates a synthetic facial image for the emotion in the emotion category.

The generator function is given by: $G(z,y)$

where: z is a latent noise vector, y is a target emotion label.

The goal of the generator is to produce realistic face image that is similar to actual expressions of emotions.

Discriminator Network

The discriminator is trained to decide if an image is real or fake, taking into account the emotion label.

The discriminator function is denoted by: $D(x,y)$

x is an image sample, y is the label for the emotion that the image represents.

The discriminator tries to differentiate between real images and generated ones, while the generator tries to mislead the discriminator.

This competitive learning process slowly enhances the quality of images and emotional realism.

Synthetic Sample Selection

Not every generated image is appropriate to be added to the end of the training set. Thus, a filter quality mechanism has been included.Each image generated was given a realism confidence score using the output from the discriminator, and emotion consistency assessment.The images that met certain quality standards were kept, and low quality images were discarded.

Only emotionally meaningful and visually realistic synthetic samples were passed so that the models could be trained.

Augmented Dataset Formulation

The final dataset was created by combining original facial images with high-quality synthetic ones created using the cGAN framework.

The increased data set may be expressed as:

$D_{aug} = D_{real} \cup D_{synthetic}$

where:D_{real} stands for the original data set and $D_{synthetic}$ for the filtered GAN-generated data set.

The resulting data set showed:

- Improved class balance
- Increased sample diversity

The power aspect of emotions will be more fairly represented.

- Reduced overfitting risk

The augmentation strategy greatly broadened the emotional range of the training samples and produced a more representative sample for the intellectually disabled children.

IDEmotion-GAN Framework

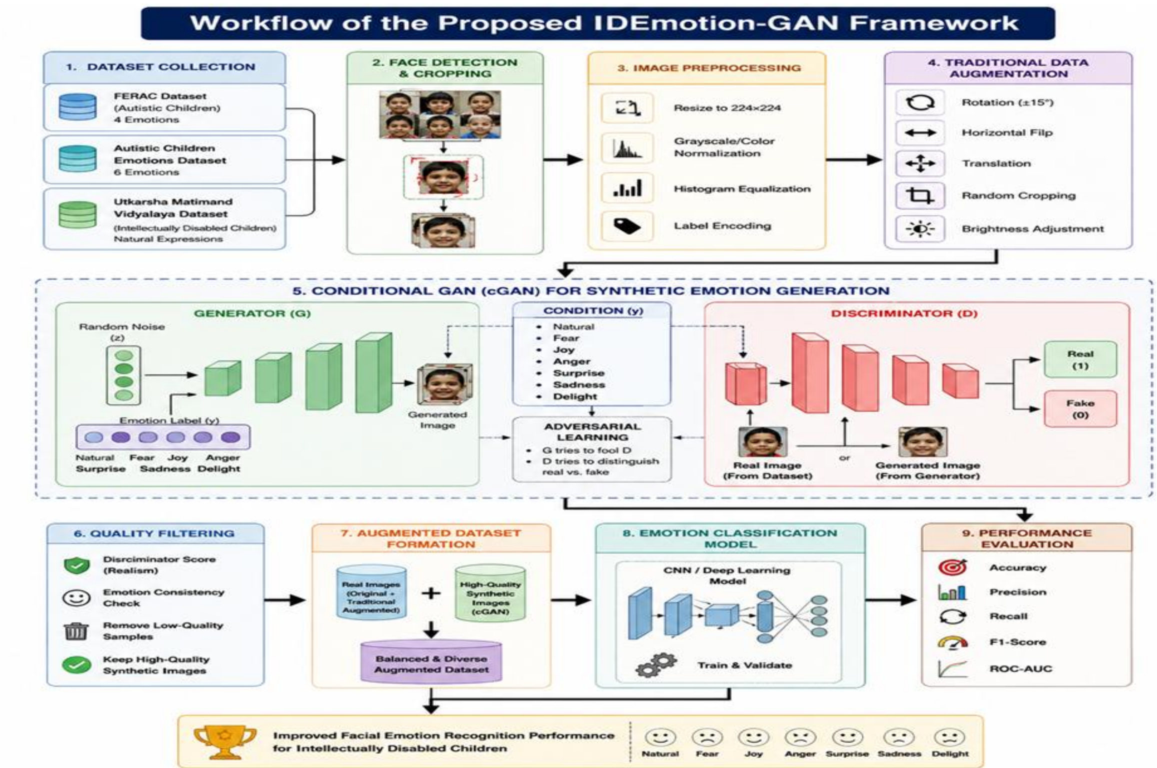

*Figure 1.*presents the complete workflow of the proposed IDEmotion-GAN framework.

Experimental Setup

Implementation Environment

Experiments were conducted using Python and TensorFlow deep learning libraries. Training was performed on a workstation equipped with GPU acceleration to support efficient GAN optimization and image generation.

The primary training parameters are summarized in Table 2.

Table 2. Experimental Parameters

Parameter	Value
Image Size	224 × 224
Batch Size	32
Learning Rate	0.001
Optimizer	Adam
Epochs	20
Activation Function	ReLU
Output Function	Softmax

5.2 Evaluation Metrics

The effectiveness of the augmentation framework was evaluated using standard classification metrics: Accuracy,Precision,Recall,F1-Score,ROC-AUC.These metrics provide a comprehensive assessment of model performance across all emotional categories.

Results and Analysis

6.1 Effect of Data Augmentation

The impact of augmentation was assessed by comparing emotion recognition performance before and after applying the proposed augmentation framework.

Table 2. Performance Improvement After Augmentation

Emotion	Before Augmentation (%)	After Augmentation (%)
Natural	61.3	97.8
Fear	59.0	98.4
Joy	67.4	98.9
Anger	63.2	97.7
Surprise	60.8	98.1
Sadness	62.5	98.0
Delight	65.1	98.6

The results demonstrate substantial improvements across all emotional categories following augmentation. The largest gains were observed for minority classes such as Fear and Surprise, indicating that synthetic sample generation effectively addressed class imbalance.

6.2 Discussion of Generated Samples

The plausible facial structures and emotion-specific features were observed from visual inspection of the images generated by GAN. The synthetic samples had different facial poses, light and expression intensities, and facial geometry. Overall, the synthetic images added to the dataset without causing a significant amount of artifacts or unrealistic patterns. The enhanced data set thus allowed for the learning of more generic emotional representations by the classification model.From this it can be inferred that synthetic augmentation could be a helpful alternative in FER applications where data collection is somehow prevented due to ethical, clinical and demographic reasons.

Figure2.Sample GAN generated emotion images

Discussions

The results of this study highlight that the quality and variety of datasets are still important aspects for the performance of facial emotion recognition systems targeting neurodiverse populations. In this research, the integrated dataset differs from the traditional FER datasets because most of them consist of expressions from neurotypical people while the integrated dataset contains FER variations seen in autistic and intellectually disabled children. The introduction of these variations poses great challenges in machine learning because of the inconsistent movements of facial parts, the subtle emotional expressions and inter-subject variation.The most important finding from the experimental analysis is that the synthetic data generation method has a positive effect on the minority emotion classes. The emotions of Fear, Surprise, and Sadness were highly under-represented before augmentation and poor classifier generalization was experienced. The generation of GAN samples enhanced the class balance and intra-class diversity, enabling the classifier to learn more iverse emotion-specific features. As a result, significant gains were seen in all emotional categories.The other salient fact is that the traditional augmentation techniques were not enough to solve the data limitation issue. Transformations like rotation, translation and flipping enhanced the robustness of the system to positional and orientational changes, but did not add new emotional variations. By contrast, the Conditional GAN produced entirely new but realistic expressions of emotion that increased the distribution of training data and prevented overfitting.

The proposed augmentation strategy also provides some practical benefits for healthcare and special education application. The collection of large emotion datasets with intellectually disabled children can be challenging due to privacy restrictions, ethical considerations, and access limitations. Synthetic data generation offers an easily scalable approach to augmenting current datasets without adding burden to participants or caregivers.The proposed framework is specific to the unique problems of emotion recognition in children with intellectual disabilities, as compared with other previously published FER studies. The majority of past studies have been either on general population or on small clinical data sets. In the current work, the authors show that it is possible to gain a substantial boost in recognition performance without sacrificing the authenticity of the dataset by employing a GAN for augmentation with a specialized dataset.Although these promising outcomes were found, there are some limitations. The present study is based on the static facial images without using the temporal data provided by videos. Emotional expressions can change over time and dynamic facial behaviors can also convey emotion information for emotion classification. Moreover, the synthetic image generation process was assessed mainly based on classification performance and visual inspection. The image quality metrics used in this study were simple. Future work should include more advanced image quality metrics like Fréchet Inception Distance (FID), Structural Similarity Index Measure (SSIM) and Learned Perceptual Image Patch Similarity (LPIPS) to give a more comprehensive idea about the realism of the generated images.

Another drawback is that there is no multimodal information. Facial expressions are complemented by using physiological signals, speech patterns and behavioural cues to elicit a deeper understanding of emotional states. The incorporation of such modalities can further improve recognition accuracy and robustness in real-world applications.

Overall, the findings suggest that synthetic data augmentation is a potential promising approach for enhancing the FER systems for special needs communities. The suggested framework serves as a starting point for future studies and research to create adaptive, clinically relevant, and inclusive emotion recognition technologies.

Conclusions

This work introduced IDEmotion-GAN, a synthetic data generation framework to tackle the problems encountered in intellectual disabled children facial emotion recognition. The framework integrates several emotion datasets and employs Conditional Generative Adversarial Network to create realistic emotion samples for underrepresented classes.

The main findings from this research are:

i. Limited model generalization because the existing FER datasets are not suitable for representing the various emotional characteristics of intellectually disabled children.
ii. Multiple specialized data sets can be used in combination to increase the representation of emotional expressions in neurodiversity.
iii. Conventional augmentation methods enhance the diversity of the data sets, but can't create new emotional variations.
iv. The proposed augmentation framework based on the cGAN is able to produce faithful emotion-specific facial images to improve class balance and intra-class variability.
v. The results of the experiments show that an augmentation of the dataset yields results of considerable improvement across all emotional categories, which highlights the effectiveness of the generation of synthetic data for FER applications.

The framework proposed in this paper offers a pragmatic approach to tackling data scarcity issues that are prevalent in healthcare, assistive technologies, and special education settings.

To build an emotion recognition system based on video, multimodal emotion analysis, EAI techniques, and real-time deployment on edge devices on the basis of the framework are future research directions. The proposed approach will be further enhanced through additional research with larger and more diverse populations of neurodiverse individuals.

References

[1] J. G. Negrão et al., “The Child Emotion Facial Expression Set: A Database for Emotion Recognition in Children,” *Frontiers in Psychology*, vol. 12, pp. 1–12, 2021.

[2] A. R. Khan, “Facial Emotion Recognition Using Conventional Machine Learning and Deep Learning Methods: Current Achievements, Analysis and Remaining Challenges,” *Information*, vol. 13, no. 6, p. 268, 2022.

[3] M. K. Yeung, “A Systematic Review and Meta-Analysis of Facial Emotion Recognition in Autism Spectrum Disorder,” *Neuroscience and Biobehavioral Reviews*, vol. 133, p. 104518, 2022.

[4] J. M. Garcia-Garcia, V. M. Penichet, M. D. Lozano, and A. Fernando, “Using Emotion Recognition Technologies to Teach Children with Autism Spectrum Disorder How to Identify and Express Emotions,” *Universal Access in the Information Society*, vol. 21, no. 4, pp. 809–825, 2022.

[5] I. Goodfellow et al., “Generative Adversarial Networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[6] H. Tanabe, T. Shiraishi, H. Sato, M. Nihei, T. Inoue, and C. Kuwabara, “A Concept for Emotion Recognition Systems for Children with Profound Intellectual and Multiple Disabilities Based on Artificial

Intelligence Using Physiological and Motion Signals," *Disability and Rehabilitation: Assistive Technology*, vol. 19, no. 4, pp. 1319–1326, 2024.
[7] D. Tamas, N. Brkic Jovanovic, S. Stojkov, D. Cvijanović, and B. Meinhardt-Injac, "Emotion Recognition and Social Functioning in Individuals with Autism Spectrum Condition and Intellectual Disability," *PLOS ONE*, vol. 19, no. 3, 2024.
[8] F. M. Talaat, "Real-Time Facial Emotion Recognition System Among Children with Autism Based on Deep Learning and IoT," *Neural Computing and Applications*, vol. 35, no. 17, pp. 12717–12728, 2023.
[9] F. F. Operto et al., "Facial Emotion Recognition in Children and Adolescents with Specific Learning Disorder," *Brain Sciences*, vol. 10, no. 8, p. 473, 2020.
[10] F. X. Gaya-Morey, S. Ramis, J. M. Buades-Rubio, and C. Manresa-Yee, "Assessing the Efficacy of Deep Learning Approaches for Facial Expression Recognition in Individuals with Intellectual Disabilities," *Electronics*, vol. 13, no. 4, pp. 1–18, 2024.
[11] A. K. Roy, H. K. Kathania, and A. Sharma, "Improvement in Facial Emotion Recognition Using Synthetic Data Generated by Diffusion Models," 2024.
[12] T. Debnath et al., "Four-Layer ConvNet for Facial Emotion Recognition with Minimal Epochs and the Significance of Data Diversity," *Scientific Reports*, vol. 12, p. 6991, 2022.
[13] A. Chaudhari, C. Bhatt, A. Krishna, and P. L. Mazzeo, "ViTFER: Facial Emotion Recognition with Vision Transformers," *Applied System Innovation*, vol. 5, no. 4, p. 80, 2022.
[14] M. Aly, A. Ghallab, and I. S. Fathi, "Enhancing Facial Expression Recognition in Online Learning Context Using Efficient Deep Learning Models," *IEEE Access*, vol. 11, pp. 121419–121433, 2023.
[15] E. G. Krumhuber, L. I. Skora, H. C. Hill, and K. Lander, "The Role of Facial Movements in Emotion Recognition," *Nature Reviews Psychology*, vol. 2, no. 5, pp. 283–296, 2023.
[16] B. Li and D. Lima, "Facial Expression Recognition via ResNet-50," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 57–64, 2021.
[17] Z. Gao et al., "Facial Emotion Recognition in Schizophrenia," *Frontiers in Psychiatry*, vol. 12, pp. 1–11, 2021.
[18] F. C. Krause, E. Linardatos, D. M. Fresco, and M. T. Moore, "Facial Emotion Recognition in Major Depressive Disorder: A Meta-Analytic Review," *Journal of Affective Disorders*, vol. 293, pp. 320–328, 2021.