Autism Rehabilitation with Meltdown Prediction Using Facial Expression and Hand Gesture Detection

by

Examination Roll:
Afroja Ahmed Smrity - 192304
Md Ziaur Rahman Joy - 192329
Sayd Mahfid Rahman - 192331

A Researech proposal submitted to the
Institute of Information Technology
in partial fulfillment of the requirements for the degree of
Bachelor of Science in
Information and Communication Technology

Supervisor: Prof. Dr. M Shamim Kaiser

Institute of Information Technology
Jahangirnagar University
Savar, Dhaka-1342

Novembar 2023

DECLARATION

We hereby declare that this thesis is based on the results found by ourselves. Materials of work found by other researchers are mentioned by reference. This thesis, neither in whole nor in part, has been previously submitted for any degree.

Afroza Ahmed Srity	Md Ziaur Rahman Joy	Sayd Mahfid Rahman
Roll: 192304	Roll: 192329	Roll: 192331

CERTIFICATE

The project titled **Metaverse Rehabilitation: Physical and Mental Rehabilitation for Pediatric Patients** submitted by **Afroza Ahmed Srity, ID: 192304**, **Md Ziaur Rahman, ID: 192329**, **Sayd Mahfid Rahman, ID: 192331** has been accepted and approved in partial fulfilment of the requirement for the degree Bachelor of Science (hons.) in Information and Communication Technology.

Prof. Dr. M. Shamim Kaiser
Supervisor

BOARD OF EXAMINERS

Prof. Dr. Fahima Tabassum
Chairman

Prof. Dr. Mohammad Shahidul Islam
Member

Prof. Dr. M Mesbahuddin Sarker
Member

Prof. Dr. Md. Hasanul Kabir
External Member

ACKNOWLEDGEMENTS

We feel pleased to have the opportunity of expressing our heartfelt thanks and gratitude to those who all rendered their cooperation in making this report.

This thesis is performed under the supervision of Dr. M. Shamim Kaiser,Professor, Institute of Information Technology (IIT), Jahangirnagar University, Savar, Dhaka. During the work, he has supplied us a number of books, journals, and materials related to the present investigation. Without his help, kind support and generous time spans he has given, we could not perform the thesis work successfully in due time. First and foremost, we wish to acknowledge our profound and sincere gratitude to him for his guidance, valuable suggestions, encouragement and cordial cooperation.

We express our utmost gratitude to Dr. M. Shamim Kaiser, Director, IIT, Jahangirnagar University, Savar, Dhaka, for his valuable advice that have encouraged us to complete the work within the time frame. Moreover, we would also like to thank the other faculty members of IIT who have helped us directly or indirectly by providing their valuable support in completing this work.

We express our gratitude to all other sources from where we have found help. We are indebted to those who have helped us directly or indirectly in completing this work.

Last but not least, we would like to thank all the staff of IIT, Jahangirnagar University and our friends who have helped us by giving their encouragement and cooperation throughout the work.

ABSTRACT

The introduction of Metaverse technology into healthcare rehabilitation, especially through web-based training, has the potential to tackle critical healthcare challenges. This research investigates how Metaverse-based solutions can transform physical and mental recovery by improving access to specialized healthcare, reducing mental health stigma, and tailoring rehabilitation programs. Utilizing machine learning to analyze eye movements and facial expressions enables early disease detection, facilitating personalized rehabilitation recommendations. This research holds both theoretical and practical significance, enhancing healthcare accessibility and patient engagement. Recommendations for future research include studying long-term effects, enhancing the user experience, addressing privacy and security concerns and global implementation strategies to harness the full potential of Metaverse-based healthcare rehabilitation on a global scale.

The use of website-based rehabilitation technology in modern healthcare is revolutionary. Patients use a user-friendly website to access rehabilitation programs while having their facial expressions and eye movements tracked for health insights. Through data analysis, potential health issues can be detected early. This innovative approach allows for tailored rehabilitation recommendations or confirmation of recovery, greatly improving patient care and outcomes.

Integrating Metaverse technology in healthcare rehab offers new avenues for physical and mental recovery, addressing access, stigma, and healthcare worker shortages. Analyzing eye movements and expressions enables early disease detection, personalized treatments, revolutionizing patient care, reducing stigma. Future research should assess long-term impact, improve user experience, ensure privacy, establish policies, and adapt to evolving tech for global use.

Keywords: Sensor, Metaverse, Rehabilitation, Physical.

LIST OF ABBREVIATIONS

ASD	Autism Spectrum Disorder
TD	Typical Development
ANN	Artificial Neural Network
FFNN	Feed Forward Neural Network
CNN	Convolutional Neural Network
LBP	Local Binary Pattern
GLCM	Grey Level Co-occurrence Matrix
SVM	Support Vector Machine
DCNN	Deep Convolutional Neural Network
VGG	Visual Geometry Group
SES	Socioeconomic status
SLD	Specific Learning Disorder
ADHD	Attention-Deficit/Hyperactivity Disorder
IVA	Integrated visual and auditory test

LIST OF FIGURES

Figure

LIST OF TABLES

Table

TABLE OF CONTENTS

CHAPTER I

Introduction

1.1 Chapter Overview

In this chapter, we present the background information, outline the problem we are addressing, highlight the scope of our study, and objectives of the research work. The Impact of the Study(National and International) and the thesis outline are discussed at the end of this chapter.

1.2 Background

Autism Spectrum Disorder (ASD) is a neuro developmental condition marked by difficulties in social communication and interaction. It affects approximately 2% of the global population and has a wide range of impacts on individuals and their families. The severity of symptoms can vary significantly, affecting different body systems and making management challenging. One particularly difficult aspect of ASD is the occurrence of acute meltdowns, which can be violent and pose risks to the individuals with ASD, their caregivers, and those around them. Recognizing and managing the behaviors leading up to these meltdowns is crucial for effective management. [1]

Studies have shown that certain behaviors, such as self-aggression, shouting, and hyperactivity, can be predictive of impending meltdowns, with speech and facial expressions characterized by high velocities and accelerations being identified as key indicators in 74.7% of cases. However, some individuals with ASD are nonverbal and struggle to express themselves through facial gestures, making it essential to utilize other forms of communication, such as gestures and sounds. Research has demonstrated promising results in detecting ASD using machine learning algorithms and computer vision tools, but there is a lack of studies specifically focused on predicting meltdowns using facial and gesture analysis. Current efforts are exploring the use of biosignals (e.g., heart rate, skin conductance) and audio signals to predict behaviors that precede meltdowns, but further investigation into facial and gesture analysis is needed. [2]

1.3 Problem Statement

The field of healthcare rehabilitation faces several critical challenges, including limited access to specialized services, persistent mental health stigma, and the need for personalized rehabilitation programs. Healthcare workers often contend with shortages, and early disease diagnosis remains a significant concern. To address these issues, we are developing a system that integrates deep learning to analyze hand movements and facial expressions. Our goal is to leverage this technology to revolutionize healthcare rehabilitation by improving accessibility, personalized care, and early disease diagnosis, while mitigating mental health stigma and enhancing patient outcomes. This research aims to provide innovative solutions to these pressing healthcare challenges.

1.4 Objectives

The presented work has been guided by the following objectives:

I) To determine if the patient has autism by using a machine learning model to analyze responses to a series of questions.

II) Proposing a real-time emotion identification system for autistic patients..

III) A total of six facial emotions are detected by the propound system: anger, fear, joy, natural, sadness, and surprise.

IV) To develop and validate a deep learning-based system that accurately predicts meltdowns in individuals with ASD by analyzing their facial and hand movements, followed by providing appropriate rehabilitation based on the predicted meltdowns.

1.5 Impact of the Study

The development and application of a machine learning-based system for predicting meltdowns in individuals with Autism Spectrum Disorder (ASD) have significant implications both nationally and internationally. This study addresses critical healthcare challenges and explores innovative solutions to improve the quality of care for individuals with ASD.

1.5.1 National Significance

In countries like Bangladesh and others with similar challenges, this study holds substantial importance for enhancing specialized care for individuals with ASD, especially in remote and underserved areas. Key impacts include:

(ii)Improving Access to Care: The proposed system, can provide real-time monitoring and early detection of meltdowns, bridging the gap for those who cannot easily access specialized healthcare facilities.

(iii)Encouraging Mental Healthcare Utilization: By offering a non-intrusive and effective way to manage and predict meltdowns, this research can help reduce the stigma associated with mental healthcare, encouraging more individuals to seek needed services.

(iv)Cost-Effective Solutions: The study explores affordable ways to implement advanced technologies in healthcare, addressing financial constraints within the national healthcare system. This can lead to broader adoption and integration of such technologies in routine care.

(v)Enhancing Caregiver Support: Providing caregivers with timely and accurate information about potential meltdowns can significantly improve the quality of care and support they offer to individuals with ASD, thereby enhancing overall patient well-being.

1.5.2 International Significance

International Significance The implications of this research extend beyond national borders, contributing to global efforts to improve healthcare access and quality for individuals with ASD. Key impacts include:

(i)Reducing Global Inequalities: By leveraging advanced technologies, the system can provide equitable access to specialized care and information, particularly in regions with limited healthcare resources. This aligns with Sustainable Development Goal (SDG) 10, which aims to reduce inequalities within and among countries.

(ii)Fostering International Collaboration: The development and implementation of this system encourage collaboration among researchers, healthcare professionals, and technology experts worldwide. Such partnerships can lead to the creation of more effective tools and strategies for managing ASD and other neurodevelopmental disorders.

(iii) Advancing Research and Innovation: This study contributes to the global body of knowledge on ASD management, particularly in the early detection of meltdowns through facial and gesture analysis. It sets a precedent for further research and innovation in the field, driving progress in healthcare technology.

1.6 Scope of the Study

We are developing a deep learning-based system to predict meltdowns in individuals with Autism Spectrum Disorder (ASD), using advanced facial and hand gesture detection technologies for real-time monitoring. By training models on diverse datasets and employing Convolutional Neural Networks (CNNs) enhancing care by providing timely insights to caregivers and healthcare professionals, especially in remote areas. It also contributes to global ASD research, promoting international collaboration and aligning with Sustainable Development Goal (SDG) 10 to reduce inequalities. However, the system cannot completely eliminate meltdowns due to the inherent variability in human behavior and ASD's complexity. It may not be universally effective due to cultural and environmental differences and is designed to assist, not replace, caregivers. Additionally, it does not address all aspects of ASD, such as cognitive, sensory, and emotional challenges, necessitating comprehensive and multidisciplinary approaches.

1.7 Research Outline

The remaining sections of the report are organized as follows: In **Chapter II**, a review of literature provides a comprehensive summary of earlier research. **Chapter III** introduces background tools and technologies.In **Chapter IV**, both the architecture of the system and the various strategies are discussed. In **Chapter V**, we conduct an in-depth analysis and discussion of result and discussion, and in **Chapter VI**, we reach a conclusion based on those findings.

CHAPTER II

Literature Review

2.1 Chapter Overview

In this chapter, the existing works have been discussed. Relative works with their limitations are also contained in this section.

2.2 Summary of the Relevant Literature

2.2.1 Works on Autism

Akhter et al [3] conducted on an analytic application of a method to justify the best Intervention Technology treating Autism-related difficulties. Based on the learning obtained from the training data, a model was built that was capable of efficiently identifying and detecting treatment technologies for a given deficit or collection of deficiencies. The first model Decision Tree Classifier had 67% accuracy and the second model named the Ensemble Vote Classifier, contained 75% of accuracy. As a result, it was applicable to autistic people's medical, nursing, mentoring, and daily requirements.The findings of the study can be used to compare and contrast survey-based research trends in the field of results, users, and disorders targeted.

In another study, Altay et al [4] did a study where kids aged 4 to 11 years were tested to be analyzed by the ASD disorder classification approach. The LDA and the KNN algorithms were applied in the classification approach. As a result of the application, the LDA algorithm provides a better result than the KNN algorithm at the accuracy rate.

Dutta et al [5] diagnosed a method that takes contributions from numerous sensors and cell phones.The study included sensor data that was brought through an application running on a portable hub. In Autism Disease diagnosis, the method showed that the most elevated exactness is 70% using Machine learning Association Rule with minimum Redundancy Maximum Relevance (mRMR) strategy.

In a brief study on ASD, Tamilarasi et al [6] initiated a deep learning-based system with ResNet-50 design and executed the examination for face image classification of ASD.The proficient deep learning strategy additionally distinguished the specific expected opportunity to erase the highlights for each image from the data put away.

Zhou et al [7] tested machine learning models to identify ASD established on multiparametric MRI data. To lessen classification error in the study of estimation bias, Cross validation was operated. These discoveries suggested that machine learning methods of multi-parametric MRI data can be advantageous in the classifying of ASD. Moreover, Particular characteristics were connected with conclusions of behavioral analysis that can also demonstrate effectiveness in keeping an eye on symptoms covering time.

2.2.2 Works on Facial Recognition

A research presented by Fatma M. Talaat describes a Deep Convolutional Neural Network (DCNN) architecture for facial expression recognition.A real-time emotion identification system for autistic youngsters was developed in this study. Face identification, facial feature extraction, and feature categorization are the three stages of emotion recognition. A total of six facial emotions are detected by the propound system: anger, fear, joy, natural, sadness, and surprise [8].

The findings by Sebastian Ludyga highlights the interplay of the physiological state and face recognition, suggesting that a single exercise induces a temporary impairment that may translate into a temporary disadvantage in social communication [9].

Researcher Salma Kammoun Jarraya and Marwa Masmoudi proposed two novel approaches for recognizing micro-expressions in autistic children during Meltdown crises .One approach focused on geometric features, while the other extracted deep features using pre-trained models and supervised learning techniques. These approaches were used to develop an "Autistic Meltdown Detector" (AMD) to ensure the safety of autistic children. Experimental results demonstrated high performance in both quantitative and qualitative evaluations, outperforming existing methods [10].

A study by Morched Derbali built a deep learning web app to diagnose autism using a convolutional neural network and camera footage of a youngster playing a video game. According to the study CNN's architecture extracted facial attributes by generating facial feature patterns and assessing facial landmark distances, classifying faces as autistic or not. VGG CNN Model produced accurate results [11].

2.2.3 Works on Meltdown

A study conducted by Rane et al [12] proposed a Logistic Regression model in order to create an Open Source based classifier for predicting ASD meltdowns. 539 samples were collected from ABIDE database. Later, LASSO regression and 5-fold cross-validation were used basically to get rid of over-fitting in the final stage, achieving an accuracy of 62%.

Hyde et al [13] used a Decision tree to examine a new way for predicting ASD with the help of employment power. The proposed model was successfully able to figure out the ASD rate within all hired employers and it specifically showed an accuracy rate of 75% and 82% that shows a high hope in predicting ASD meltdowns.

Gong et al [14] continued his existing work with the help of Association rules. The author drew out genes associated with ASD from 16,869 publication abstracts. With both candidate genes and identified genes, Association rules were applied. This research also highlights the potential of text classification for predicting undiscovered genes in ASD meltdowns in the future.

2.2.4 Gaps in the Literature

As per the literature review we can clearly conclude the limitations we found during our observation.Here we will talk about some drawbacks we found in the related works in table 2.1

Table 2.1: **Limitation of Existing Works**

Reference	Author	Based on	Approach	Year	Limitations
[15]	Kyoung-Hwan Cho, Jeong-Beom Park	Exercise Rehabilitation	Delphi technique	2023	A total of twenty experts assisted the whole exercise, which is sometimes very difficult to achieve.

Table 2.1 – **Limitation Of Existing Works(Continued)**

Reference	Author	Based on	Approach	Year	Limitations
[16]	Ibrahim Abdulrab Ahmed, Ebrahim Mohammed Sena, Taha H. Rassem	Autism	Machine Learning and Deep Learning	2022	ANN algorithm failed by 0.3% to classify one image as TD and classified it as ASD. Whereas FFNN, GoogLeNet, ResNet-18 failed by 0.3%, 6.4%, 2.4% respectively.
[17]	Zhong Zhao, Haiming Tang, Xiaobin Zhang	Autism using Eye Tracking data from Face to Face Conversation	Machine Learning	2021	Individuals with ASD are more sensitive to wearing devices. This study utilized a head-mounted eye tracker to record the gaze behavior, which might affect the social behavior of children with ASD to a larger extent.
[8]	Fatma M. Talaat	Autism	Deep Learning and IoT	2023	Lackings of synchronization in Cloud, Fog, and IoT layers may cause trouble. Photograph location needs special attention.
[9]	Sebastian Ludyga, Markus Gerber, Fabienne Bruggisser	ASD	Behavioral Assessments	2023	(i) The inclusion of only one girl limits the generalizability of our results. (ii) Did not consider the influence of the individual comorbidity profile due to the small sample size.

Table 2.1 – **Limitation Of Existing Works(Continued)**

Reference	Author	Based on	Approach	Year	Limitations
[18]	Sima Aminoleslami, Keivan Maghooli	Autism	Neural Network	2023	The system is specially designed for children. Not applicable for adults due to differences in the training parameters of the system.
[19]	Agnese Capodieci, Marco Romano	Language and Learning Disorders	MemoRAN	2022	(i)small sample size. (ii) no information about SES and parental education was presented.
[20]	Aliakbar Pahlevanian,Nader Alirezaloo	ADHD	Neurofeedback (NF) ,Cognitive training (CT)	2015	Neurofeedback and cognitive rehabilitation therapy can only be applied in the attention and impulsion areas of response inhibition of children with ADHD and not to other areas of response inhibition or other populations.
[21]	Lisa Feldman Barrett,Batja Mesquita, Kevin N.Ochsner, James J.Gross	Emotion Recognition	Facial action coding system (FACS)	2007	The algorithm has trouble categorizing facial expressions and is unable to recognize and detect any underlying emotions that a viewer may feel when they view the image
[22]	Derick Leony a, Pedro J. Muñoz-Merino a, Abelardo Pardo b	Emotion Recognition	Visualizations In a computer interaction based environment	2013	Small dataset (limited scale)

Table 2.1 – **Limitation Of Existing Works(Continued)**

Reference	Author	Based on	Approach	Year	Limitations
[23]	Magali Batty, Margot J. Taylor	Facial Emotional expression	Intensity (10%, 55%, 90%); kind of expression (anger, disgust, fear, joyful, sad, surprise); and observer gender as a subject factor	2016	Neutral replies that were 90% neutral and 10% expressive for extremely low intensity expressions were recorded as erroneous
[24]	Md Inzamam Ul Haque; Damian Valles	Facial Expression recognition on autistic children	CNN model	2018	Uncontrolled environment
[25]	Sara Karim; Nazina Akter; Muhammed J. A. Patwary	ASD Meltdown	Fuzzy semi-supervised learning	2022	Mid-fuzziness segments were more likely to misclassified ASD Meltdown

CHAPTER III

Background Tools And Technologies

3.1 Overview

In this chapter, an overview of numerous commonly used neural networks and the related terminology are presented, as well as an introduction to the libraries that are used in the building of networks. The purpose of this part is to acquaint the readers with the many resources and ideas that will be referred to and employed over the whole of the book.

3.2 Machine Intelligence Libraries

The models were built using code taken from a number of different libraries; we will go through some of these libraries in the next section.

3.2.1 TensorFlow

TensorFlow is a popular open-source machine learning framework developed by Google. It is designed to support a wide range of tasks, including image and speech recognition, natural language processing, and time-series analysis. TensorFlow uses a dataflow graph to represent mathematical computations, making it easy to parallelize and distribute computations across multiple CPUs or GPUs. The framework also includes a range of tools for data preprocessing, model selection, and evaluation, as well as a powerful and flexible API for building and training deep neural networks. With its scalable architecture and extensive functionality, TensorFlow has become a widely used tool for researchers, developers, and businesses across a range of industries.

3.2.2 Keras

Keras is a high-level deep learning framework that provides a user-friendly interface for building and training deep neural networks. It was developed with the aim of making it easier for researchers and developers to experiment with deep learning models without requiring extensive knowledge of the underlying mathematical concepts. One of the key features of Keras is its modularity, which allows users to easily construct and combine various layers to form complex neural network architectures.Keras supports a wide range of layer types, including convolutional layers for image processing, recurrent layers for sequence data, and dense layers for general purpose modeling. Keras also provides a number of pre-built models for common deep learning tasks, such as image classification and language processing. These models are easily accessible and can be customized to fit specific needs, making it easy to quickly develop and test new ideas. Another strength of Keras is its support for multiple backend frameworks, including TensorFlow, Theano, and CNTK. This allows users to choose the backend that best fits their needs, while still taking advantage of Keras' high-level API.

3.2.3 Scikit Learn

Scikit Learn is a popular Python library for machine learning that provides a range of tools for data preprocessing, classification, regression, clustering, and more. It is designed to be simple and efficient, making it easy for both beginners and experienced data scientists to work with large datasets and develop models quickly. One of the strengths of Scikit Learn is its focus on consistency and usability. The library provides a uniform interface for a wide range of machine learning algorithms, allowing users to easily switch between different models and compare their performance. The library also includes a range of preprocessing tools, such as feature scaling and dimensionality reduction, which can help improve the accuracy of machine learning models.

3.2.4 CV2

The cv2 module is the main module in OpenCV that provides developers with an easy-to-use interface for working with image and video processing functions.[100]

In Python, OpenCV is accessed through the cv2 module, which provides a user-friendly interface to leverage the powerful features of the OpenCV library. With cv2, developers can easily read, manipulate, and analyze images and videos, performing tasks that range from simple operations like resizing and cropping to complex ones like object detection and image segmentation.cv2 empowers developers to create innovative solutions that leverage the full potential of computer vision technology.

3.2.5 Numpy

NumPy is a fundamental Python library for scientific computing, providing powerful support for n-dimensional arrays (ndarray), which are efficient, fixed-size, and homogeneous. It enables advanced mathematical operations, including element-wise computations, linear algebra, statistics, and random number generation, all optimized for performance with pre-compiled C code. Key features like vectorization and broadcasting simplify and speed up array operations. Widely used in data science and machine learning, NumPy is essential for efficient manipulation and analysis of large datasets.

3.2.6 Mediapipe

MediaPipe is an open-source framework for building pipelines to perform computer vision inference over arbitrary sensory data such as video or audio. Using MediaPipe, such a perception pipeline can be built as a graph of modular components.[101]

It is a versatile framework that leverages the power of GPUs for faster processing of multimedia tasks. It utilizes parallel processing to handle multiple tasks simultaneously, such as processing multiple video streams or running several computer vision models. Additionally, MediaPipe integrates with OpenCV, a robust library for computer vision, to enhance its capabilities for tasks like video capture, processing, and rendering. By teaming up with TensorFlow, Google's machine learning tool, MediaPipe simplifies the integration of pre-trained or custom models for tasks such as face recognition and speech understanding. With support for popular programming languages like C++, Java, and Python, MediaPipe is easy to incorporate into various projects, making it a valuable tool for multimedia processing and machine learning applications.

3.2.7 Scipy Spatial

Scipy Spatial Distance is a module in the Scipy library that provides functions for calculating distances between points in n-dimensional space. It also includes functions for computing distance matrices, which are matrices that contain the distances between all pairs of points in a given set.

The Scipy Spatial Distance module offers a wide range of distance metrics, including Euclidean distance, Chebyshev distance, Hamming distance, and many more. Each metric has its own mathematical formula for calculating distances between points.It also provides functions for working with data sets that have missing or invalid values. These functions can help ensure that your calculations are accurate even when dealing with imperfect data.Overall, Scipy Spatial Distance is a powerful tool for anyone working with spatial data in Python. Whether you're analyzing geographic data, clustering data points, or performing machine learning tasks, this module can help you accurately measure distances and make informed decisions based on your results.

3.3 Preprocessing Strategies

Throughout the preprocessing stages of the work we have reshaped and normalized the pixel values.Facial images typically have only one channel (grayscale). Our code adds a new dimension using np.expand dims to make it compatible with the Convolutional Neural Network (CNN) model, which expects a format like (number of images, height, width, channels).The code uses to categorical from tensorflow.keras.utils to convert the emotion labels (likely integer values representing different expressions) into one-hot encoded vectors which is a common technique for representing categorical data in neural networks, where each category has its own binary vector with a single "1" and all other values as "0".

3.4 Neural Networks Related to Our Work

In this section, a brief overview is provided of the neural networks that pertain to the study. This summary aims to familiarize the neural networks that are most relevant to the study and set the stage for the subsequent discussions on their applications and performance.

3.4.1 Convolutional Neural Network

A convolutional neural network (CNN) is a special kind of artificial neural network that was developed for the sole purpose of processing pixel input. CNNs are used in image recognition and digital processing. Convolutional neural networks make use of a variety of filters in order to identify picture characteristics that allow for the classification of objects. The extraction of these features is being handled by a great number of kernels inside the CNN system. It is common practice to call attention to noteworthy differences in pixel values by using the edge kernel. CNNs are used to translate two dimensional or multi- dimensional input to a single dimensional output variable. They have proved to be so beneficial that they are currently the technique of choice for solving any kind of prediction problem that involves picture data. The components of a convolutional neural network are as follows: an input layer, an output layer, and hidden layers. Any layers in a feedforward neural network that are located in the middle are referred to be hidden since the activation function and the final convolution both conceal the inputs and outputs of those layers. As compared to a standard neural network, a convolutional network is highly distinct due to the fact that the neurons in each of its layers are organized according to height, weight, and depth.

3.4.2 Dense

When one layer in a neural network is densely coupled to the layer above it, this means that each neuron in the layer is connected to each and every other neuron in the layer above it. This layer is the one that is used in artificial neural network networks the most often, hence it is the one with the highest prominence. In a model, each neuron in the layer below it transmits signals to the neurons in the layer above it, which are responsible for multiplying matrices and vectors. During matrix vector multiplication, the row vector of the output from the layers that came before it is the same as the column vector of the layer that contains the dense information. In order to multiply matrices with vectors, the row vector and the column vector must both have the same number of columns.

3.4.3 Flatten

It is common practice to employ a neural network's Flatten layer, which is a simple but necessary layer, to convert the output of a convolutional or pooling layer into a feature vector that is two-dimensional in shape. This layer is responsible for taking the output of a multidimensional array from the layer below it and turning it into a single vector so that it may be sent on to a fully linked layer for further processing. The Flatten layer is essential because the majority of machine learning methods, including logistic regression, need for the input data to be in the form of a flat vector. This makes the Flatten layer an essential component. Neural networks are able to readily integrate with these algorithms and take benefit of the predictive potential they provide if they make use of the Flatten layer. Moreover, the Flatten layer may assist minimize the number of parameters in the model by consolidating numerous feature maps into a single vector. This is accomplished via the use of vectorization. In general, the Flatten layer is a simple yet essential component of many different architectural designs for neural networks. It allows these networks to handle complicated input data in an effective manner and attain high levels of accuracy.

3.4.4 Convolution Operation

The convolutional operation is an essential component in the construction of convolutional neural networks (CNNs). The process entails applying a filter, which is often referred to as a kernel, to an input picture in order to extract features that are significant. The filter iteratively applies itself to the picture, calculating a dot product at each place along the way, and output a feature map at the end. The size of the input picture, the size of the filter, and the stride, which defines the distance the filter travels between each calculation, all influence the size of the feature map that is produced as the output of the algorithm. By altering the weights of the filters

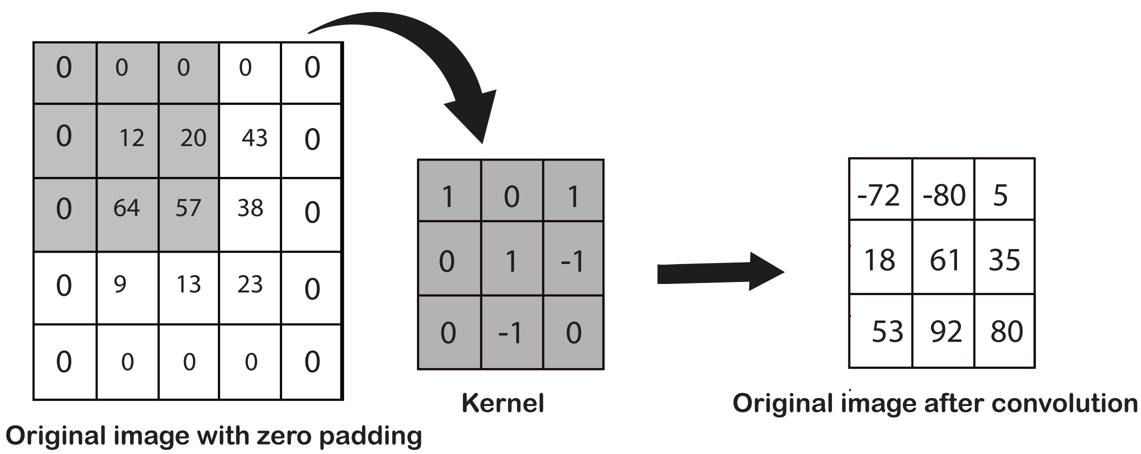

Figure 3.1: Convolution Technique

while the convolutional layer is being trained, it is possible for the layer to learn how to recognize certain characteristics, such as edges, corners, and textures. The network is able to learn hierarchical representations of the input data as a result of this,with lower levels identifying basic characteristics and higher layers identifying more complex ones. After a series of convolutional layers, a network will often go on to a series of pooling layers. These layers downsample the feature maps, hence reducing the overall size of the maps and assisting the network in becoming more generalized.

3.4.5 MaxPooling Operation

It is common practice to do maxpooling after a layer of convolutional filters has been applied. This technique is helpful for lowering the spatial dimensions of the feature maps while still preserving the characteristics that are most important to the analysis. In order to carry out the maxpooling operation, the input feature map has to be cut up into a number of pooling windows first. These windows are non overlapping rectangular regions, and they do not overlap with one another. The maxpooling technique results in an output that is the maximum value that was found to be present inside each pooling window. This output is produced by the maxpooling procedure. maxpooling greatly reduces the spatial dimensions of the feature maps by picking the feature with the highest value and pooling it with the other features. This is done while keeping the features that are most essential to the analysis.

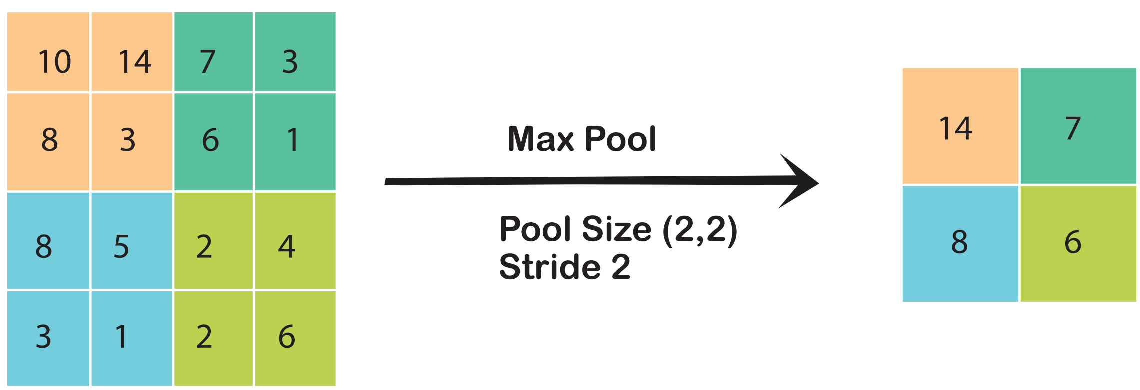

Figure 3.2: Maxpooling Techniques

3.4.6 Dropout Layer

Regularization strategies like the dropout layer are often employed in neural networks to reduce overfitting and improve accuracy. It does this by randomly removing a certain proportion of neurons from the layer that is being trained. This compels the network to acquire redundant representations and lessens the likelihood that it would become too dependent on certain characteristics. During the process of inference, all neurons are active, but the output of each neuron is scaled according to the dropout probability. This is done to maintain the constant value that is predicted from each neuron. Dropout layers may be introduced to any region of a neural network and are most successful when used in combination with large or deep networks. Dropout layers can also be removed from a neural network. They have been found to increase neural networks' generalization performance on a variety of tasks, including image recognition, natural language processing, and voice recognition, among others. The dropout approach is frequently used in the deep learning field and has developed into an important instrument for the construction of neural networks that are precise and resilient

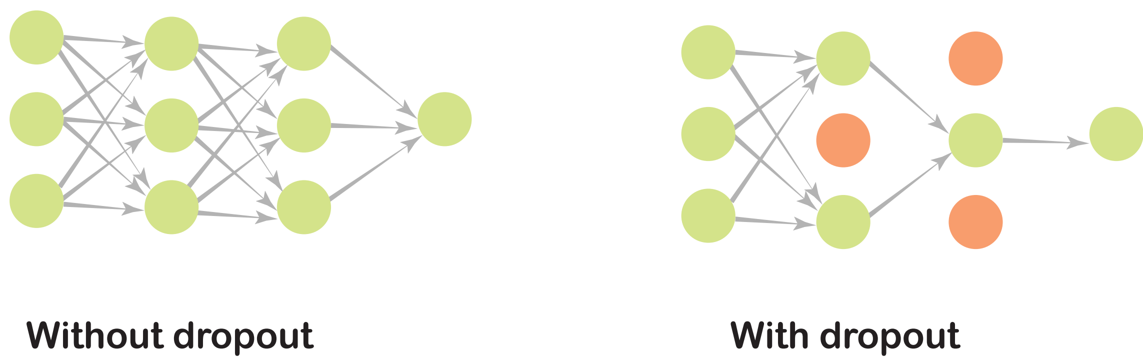

Figure 3.3: Dropout Layer

3.4.7 Activation Function

An activation function is a mathematical function that is applied to the output of each neuron in a layer of a neural network.It's job is to provide the model some non-linearity so that it may learn to recognize intricate patterns in the data. This is accomplished via the model's ability to learn. There are a wide variety of activation functions accessible, and each one has both advantages and disadvantages unique to itself. The rectified linear unit(ReLU),softmax functions are the types of activation functions that are used in our code. The ReLU function is frequently utilized in the hidden layers of a neural network due to its ease of use and its capacity to circumvent the vanishing gradient problem. On the other hand,By incorporating the softmax activation function, multi-class classification models can generate more meaningful and interpretable outputs, allowing for better decision-making and evaluation.These activation functions are all examples of what are known as "enhancement functions." The selection of an activation function is an essential factor to take into account while developing and perfecting a model since it has the potential to have a major influence on the performance and learning capabilities of a neural network.

3.4.8 Adam Optimizer

Adam optimizer is a popular optimization technique in deep neural networks that combines the advantages of Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation. Adam optimizer is used extensively in the field of artificial intelligence research (RMSProp). It is a technique for calculating individual adaptive learning rates based on the parameters, and it is referred to as an adaptive learning rate method. An ongoing estimate of the first and second moments of the gradient is kept by the Adam optimizer. This estimate is then utilized to update the parameters. This makes it a very efficient approach for improving the neural network since it helps to adapt the learning rate depending on the gradients. The capacity of the Adam optimizer to converge in a speedy and effective manner is one of the primary advantages offered by this tool. It has been shown to be very successful in a broad variety of deep learning tasks, such as the categorization of images, the processing of natural languages, and the identification of voices. The algorithm's user-friendliness and adaptability have contributed to its rise in popularity among developers and academics. These qualities enable the algorithm's users to customize the optimization process to better suit their own requirements.

3.5 Evaluation Metrics

3.5.1 Accuracy

Accuracy is a commonly used metric to evaluate the performance of a deep learning model. It measures the proportion of correctly classified instances out of the total number of instances. However, it has limitations in certain scenarios, such as when the classes are imbalanced or when the cost of misclassifying one class is much higher than the other. In these cases, other metrics such as precision, recall, and F1 score may be more appropriate. It is also important to note that accuracy alone may not provide enough information about the performance of the model and should be complemented by other metrics and visualizations such as confusion matrix, ROC curve, and precision-recall curve.

3.5.2 Recall

In deep learning, recall is a metric that measures the ability of a model to correctly identify positive instances from all actual positive instances in a dataset. It is also known as sensitivity or true positive rate (TPR). Recall is calculated as the number of true positives (TP) divided by the sum of true positives and false negatives (FN). A high recall value indicates that the model can successfully identify most of the positive instances in the dataset. However, high recall may also come at the cost of lower precision, as the model may incorrectly classify some negative instances as positive. Therefore, the balance between precision and recall is important for evaluating the overall performance of a deep learning model.

3.5.3 Precision

Precision is a performance metric in deep learning that measures the proportion of true positives among the predicted positive values. In other words, it tells us how accurate the model is when it predicts that a given sample belongs to a particular class. A high precision means that the model has a low false positive rate, which is important in tasks where a false positive can have serious consequences. For example, in a medical diagnosis, a high precision means that a patient diagnosed with a disease is more likely to have the disease. However, it is worth noting that high precision does not necessarily mean high accuracy, as a model may be precise but still miss a large number of true positive cases.

3.6 Conclusion

This chapter provides an overview of the libraries that were employed, as well as the neural networks that were pertinent to our work and the terms that were linked with them.

CHAPTER IV

SYSTEM ARCHITECTURE AND METHODOLOGY

4.1 Chapter Overview

This chapter thoroughly reviews the datasets used in the study, detailing the steps taken to prepare the data, including preprocessing functions and dataset augmentation. Additionally, it discusses the model architectures employed for classification and provides a concise explanation of how the model operates in segmentation and classification tasks.

4.2 Datasets

This study employed two distinct datasets, referred to throughout this article as Dataset 1 and Dataset 2. By using multiple datasets, the research aims to provide an in-depth analysis of the subject matter. The article will explore the distributions and characteristics of these datasets, offering a thorough examination to support the study's findings and conclusions.

4.2.1 Dataset 1

This dataset includes factors associated with the development of ASD in children, featuring the A10 Autism Spectrum Quotient (columns A1-A12), the Social Responsiveness Scale (scored out of 10), age in years, speech delay/language disorder, learning disorder, genetic disorders, depression, global developmental delay/intellectual disability, social/behavioral issues, anxiety disorder, sex, jaundice, and family history of ASD. The data comprises various quantities and factors characterizing ASD in children, with most columns returning binary values (0 or 1) except for the Social Responsiveness Scale.

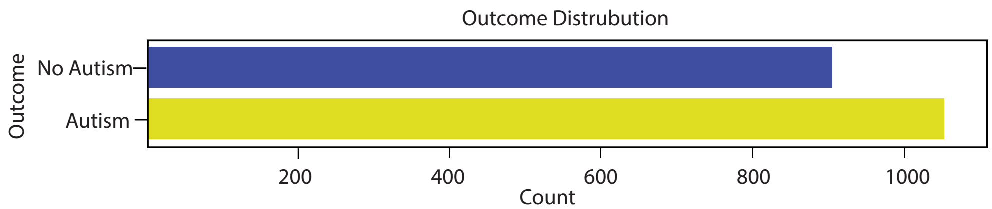

Figure 4.1: Count plot of Dataset-1

Table 4.1: **Table for data type Dataset-01**

Dataset	**ASD**	**NO ASD**	**Total**
Dataset-1	897	1065	1962

4.2.2 Dataset 2

The dataset contains 48x48 pixel grayscale images of faces, each automatically centered and uniformly scaled. The goal is to classify facial expressions into one of seven emotions: 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, and 6=Neutral. The training set comprises 28,709 images, while the public test set includes 3,589 images.

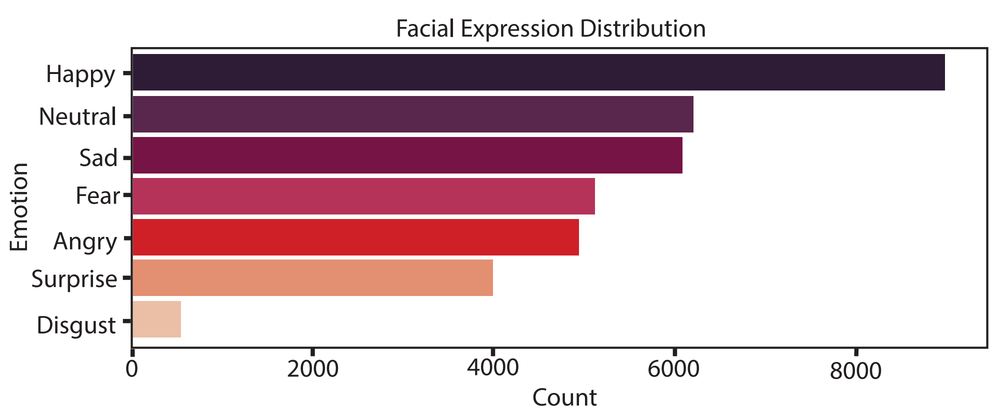

Figure 4.2: Count plot of Dataset-02

Table 4.2: **Table for data type Dataset-02**

Dataset	Happy	Neutral	Sad	Fear	Angry	Surprise	Disgust	Total
Dataset-2	8920	6330	6025	4435	4726	3897	512	34845

4.2.3 Heatmap of Dataset-01

The heatmap displays correlations in an autism dataset, revealing key predictors like family history of ASD, social responsiveness, and global developmental delay strongly linked to outcomes. Moderate correlations exist for age, learning disorders, and social/behavioral issues, highlighting their relevance in autism detection. Weaker correlations, such as jaundice and sex, suggest lesser direct impacts. This analysis guides feature selection, enhances predictive models, and improves autism diagnosis and management by prioritizing influential factors.

Figure 4.3: Heatmap(Dataset-01)

4.2.4 Proposed Architecture

The proposed architecture for the autism detection and management system involves several integrated components and processes designed to work together seamlessly to detect autism, monitor patient behavior, and suggest personalized rehabilitation strategies. The workflow begins with the collection of a diverse and comprehensive dataset aimed at training and testing an SVM (Support Vector Machine) model specifically for autism detection.

Upon detecting autism using the trained SVM model, the system triggers a background video playback while simultaneously activating the webcam with timer to monitor the patient's real-time behavior. This video could contain therapeutic scenarios or potential triggers to observe the patient's reactions. Concurrently, the system captures real-time video data through the webcam, focusing particularly on hand gestures.

In parallel, the architecture incorporates an emotion detection module. A separate dataset is collected for this purpose, containing pixel values of facial expressions labeled with corresponding emotions. This dataset undergoes several preprocessing steps, including reshape to a uniform size, extracting pixel data, and encoding emotion labels into a machine-readable format. A Convolutional Neural Network (CNN) is then trained on this preprocessed data to recognize and classify different emotions based on facial expressions, allowing the system to monitor the patient's emotional state in real-time through the webcam feed.

The integration of hand gesture detection and emotion detection modules allows the system to operate these two functionalities concurrently. By doing so, the system can provide a holistic view of the patient's state, combining physical gestures with emotional cues to identify potential meltdown points. When a meltdown is detected, the time is then synchronized with the background video's timeline to pinpoint specific scenes or events that may have triggered the meltdown.

By analyzing these triggers, the system can identify the scenarios that most affect the autism patient. Based on this analysis, the system suggests personalized rehabilitation strategies tailored to the individual's specific needs and responses. These strategies may include therapeutic interventions, environmental modifications, or other supportive measures aimed at mitigating or preventing future meltdowns.

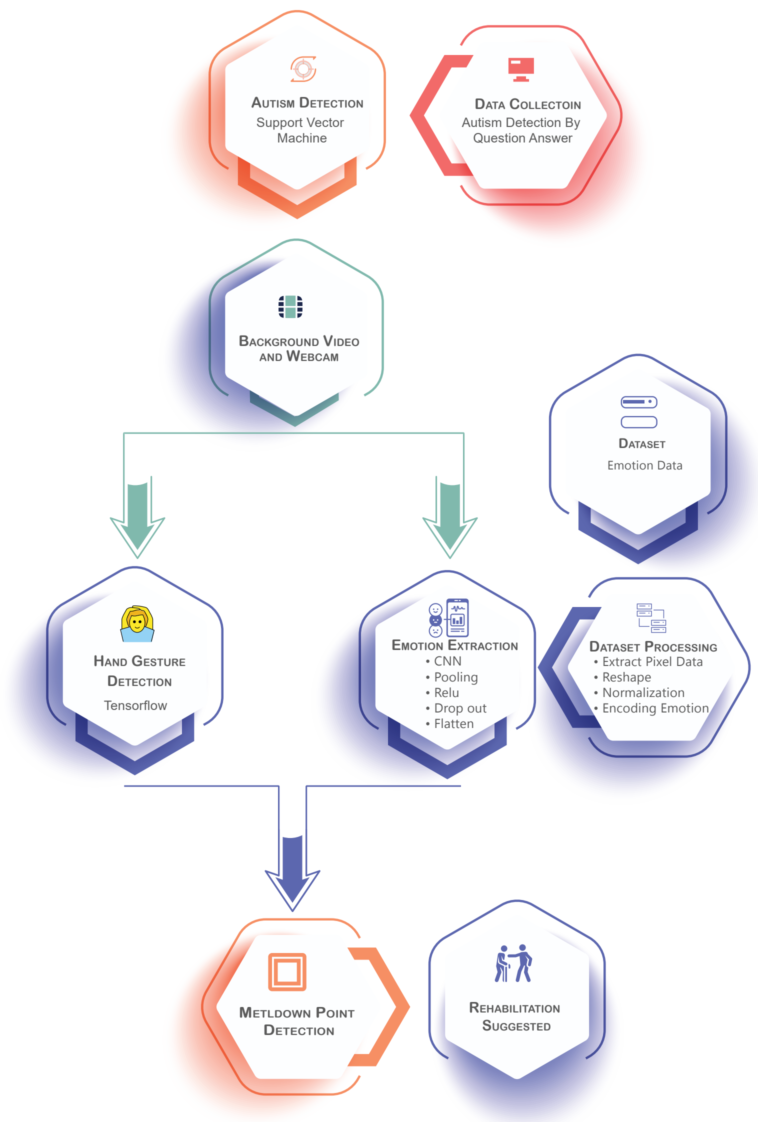

Figure 4.4: Detailed workflow of our system

4.2.5 Overview of How the Proposed Architectures Work

The proposed architectures integrate machine learning and computer vision techniques to detect autism, recognize emotions, and monitor hand gestures for detecting autism-related meltdowns. The first phase uses Python libraries like NumPy, Pandas, and Scikit-Learn to detect autism. The autism dataset is loaded into a Pandas DataFrame for exploration, splitting the data into features (X) and labels (Y), with 'Outcome' indicating autism presence. Features are standardized using 'StandardScaler' and divided into training and testing sets using 'train_test_split'. An SVM classifier with a linear kernel is trained and evaluated, with the ability to make

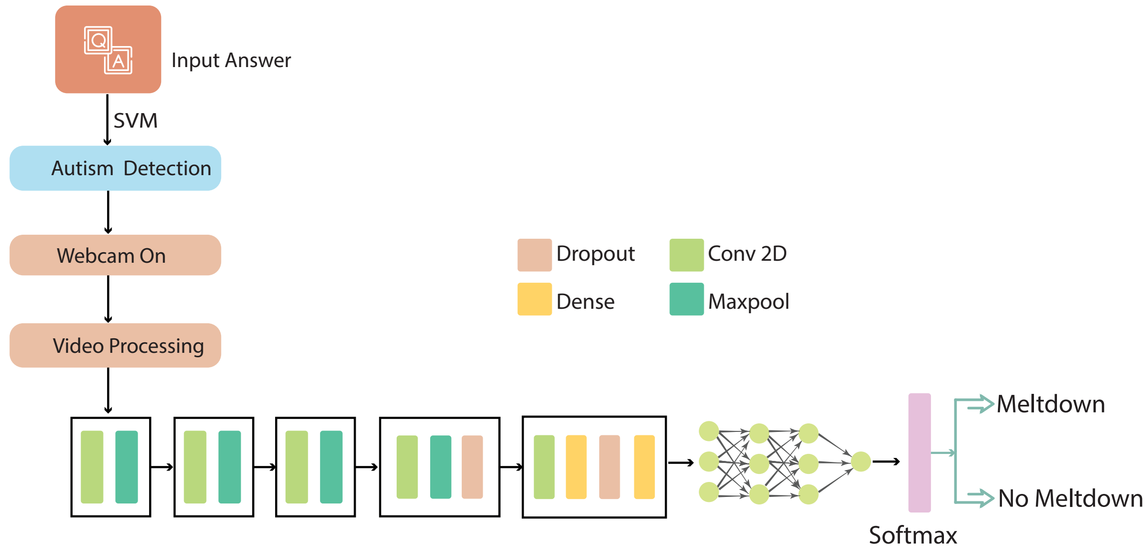

Figure 4.5: Architecture

Blocks	Name of Layers	No. of Layers	Filter Size	No. of Filters
Block 0	Conv2D	1	(3, 3)	16
	MaxPool	1	(2, 2)	
Block 1	Conv2D	1	(3, 3)	32
	MaxPool	1	(2, 2)	
Block 2	Conv2D	1	(3, 3)	64
	MaxPool	1	(2, 2)	
Block 3	Conv2D	1	(3, 3)	128
	MaxPool	1	(2, 2)	
	Dropout	1		
Block 4	Flatten	1		
	Dense	2		
	Dropout	1		

Table 4.3: **Model Architecture**

predictions on new input data to determine if an individual is likely to have Autism Spectrum Disorder (ASD).

The second phase involves building and training a Convolutional Neural Network (CNN) to recognize facial expressions from the FER-2013 dataset. The dataset is preprocessed into 48x48 grayscale images and normalized. The data is split into training and testing sets, and the CNN, comprising multiple convolutional layers, ReLU activation, max-pooling layers, and dropout layers, is compiled with the Adam optimizer and categorical cross-entropy loss function. After training for 100 epochs, the model achieves emotion classification and is saved for future use.

The final phase integrates hand gesture detection with the emotion model using OpenCV and MediaPipe to capture real-time video, detect facial landmarks, and track hand movements. The pre-trained emotion model predicts facial expressions in real-time, while MediaPipe's face mesh and hand tracking solutions identify face and hand regions. Hand coverage of the face triggers a meltdown detection mechanism, and eye direction is calculated to monitor gaze. The system overlays bounding boxes and predictions on the video feed, allowing simultaneous monitoring of hand gestures and emotional states to detect and respond to autism-related meltdowns in real-time, providing a comprehensive framework for detecting triggers and suggesting rehabilitation strategies.

CHAPTER V

Results and Discussion

5.1 Overview

This chapter includes a brief analysis of the performance of the model.

5.2 Evaluation and Analysis

This section examines the performance of the expression recognition and autism detection models through detailed evaluation metrics. By analyzing training accuracy trends and other key metrics such as F1 Score, Recall, and Precision, we aim to assess the models' effectiveness and identify areas for improvement to ensure robust and accurate autism management.

5.2.1 Expression Model Accuracy

The training accuracy of the expression model, as depicted in Figure 5.1, demonstrates a clear upward trend over 100 epochs. The initial accuracy starts around 20% and gradually increases, with notable fluctuations, reaching over 90% by the 100th epoch. These fluctuations suggest that the model undergoes significant learning and adjustment phases during training. The consistent improvement indicates that the model effectively learns from the training data, ultimately achieving high accuracy in recognizing facial expressions.

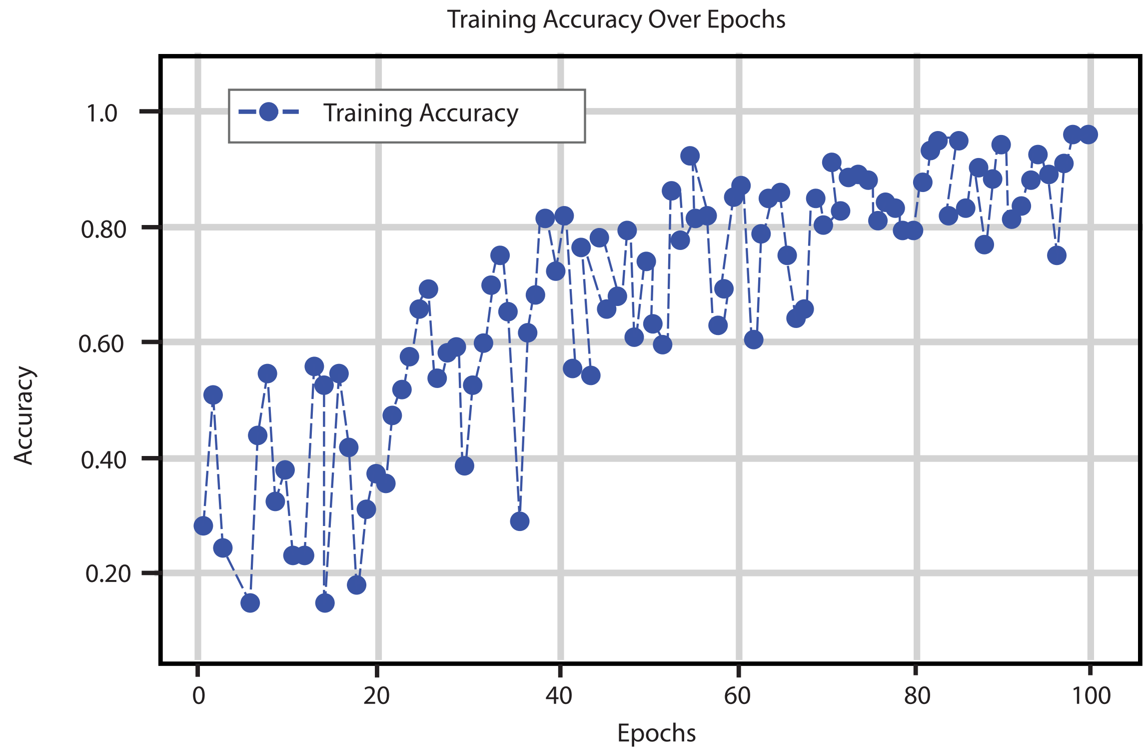

Figure 5.1: Expression Model Accuracy

5.2.2 Autism Detection Accuracy

In contrast, the SVM model for autism detection, as shown in Figure 5.2, presents a different training behavior. The accuracy initially fluctuates between 66% and 71% during the first 20 epochs but stabilizes around 69% for the remaining epochs. This plateau indicates that the model reaches its learning capacity early on and does not significantly improve with additional training. The limited improvement in accuracy might suggest that the dataset or the features used for the SVM model may require further refinement or augmentation to enhance the model's performance. Additionally, it could imply that the SVM model's complexity is insufficient to capture the nuances of the data, indicating the potential need for more advanced techniques or a different modeling approach. Furthermore, exploring additional preprocessing steps or incorporating a wider variety of features might also help in achieving better accuracy and overall performance.

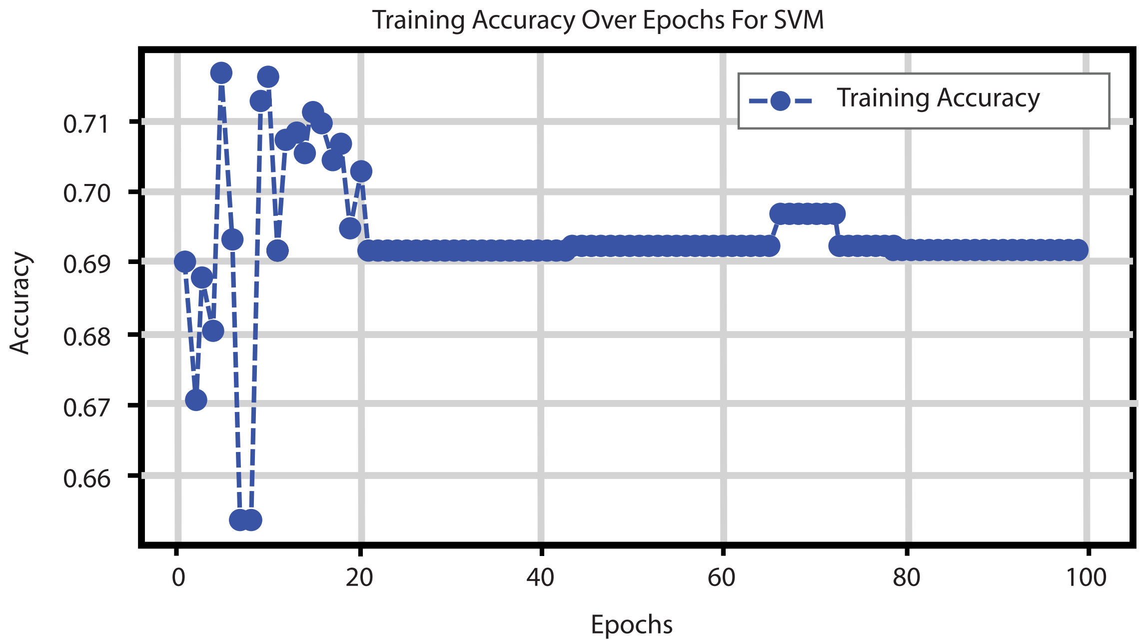

Figure 5.2: Autism Detection Accuracy

5.2.3 Evaluation Metrics for SVM

Figure 5.3 provides a comprehensive view of the SVM model's performance across multiple evaluation metrics, including F1 Score, Recall, and Precision, over 100 epochs. The training accuracy, F1 Score, and Recall exhibit similar patterns, with initial fluctuations followed by stabilization. Precision, however, shows more variability throughout the epochs but ultimately aligns with the trends observed in the other metrics.

Training Accuracy: Stabilizes around 69%, indicating the model's consistency in identifying autism-related patterns.

F1 Score: Mirrors the accuracy trend, suggesting a balanced performance in terms of precision and recall.

Recall: Similar to accuracy and F1 Score, it indicates that the model reliably identifies true positive cases.

Precision: Exhibits more variability, highlighting potential issues with false positives that might need addressing through model adjustments or data balancing.

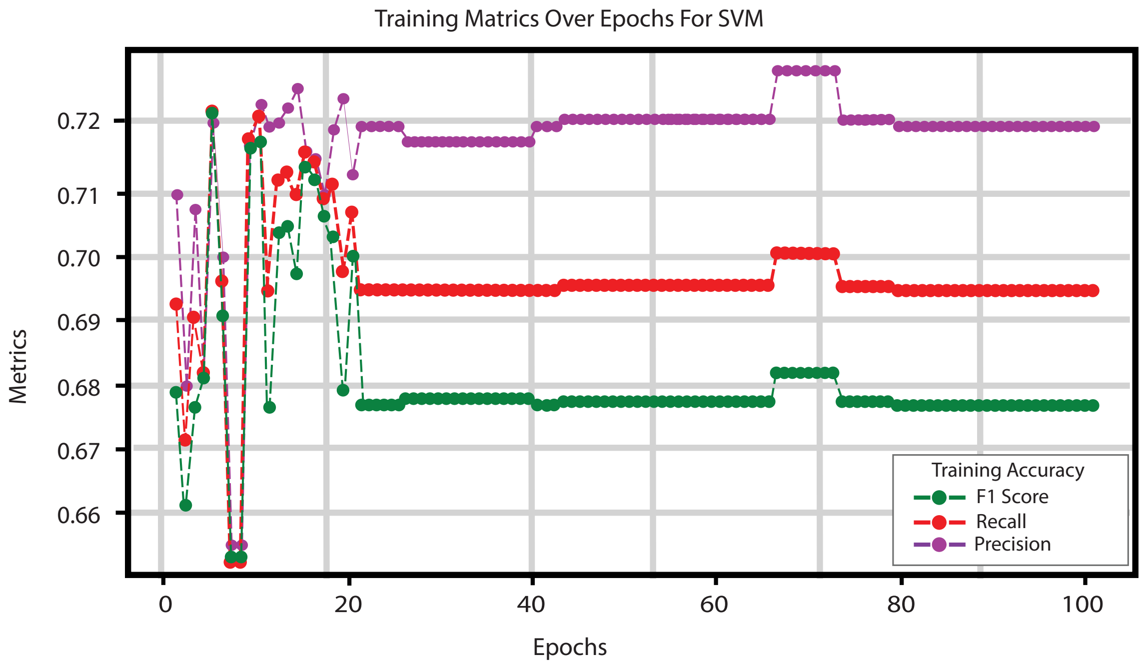

Figure 5.3: Evaluation Matrix for SVM

5.2.4 Evalutation Metrics for Expression Model

It offers a detailed overview of the expression model's performance across several key evaluation metrics, including Training Accuracy, F1 Score, Recall, and Precision, over the course of 100 epochs. The chart reveals distinct patterns and behaviors for each metric as the training progresses.

Training Accuracy: Exhibits initial fluctuations but generally stabilizes towards the later epochs, indicating the model's improving consistency in correctly identifying expressions. The final accuracy hovers around a specific value, showing the model's steady performance.

F1 Score: This metric follows a trend similar to Training Accuracy, reflecting a balanced performance concerning both precision and recall. The stabilization of the F1 Score suggests that the model maintains a good balance between precision and recall as training continues.

Recall: The recall metric also mirrors the patterns of accuracy and F1 Score, indicating that the model consistently identifies true positive cases over time. The recall stabilization aligns with the overall trend, confirming the model's reliability in recognizing true positives.

Precision: Unlike the other metrics, precision displays more variability throughout the training epochs. This indicates potential challenges with false positives, sug-

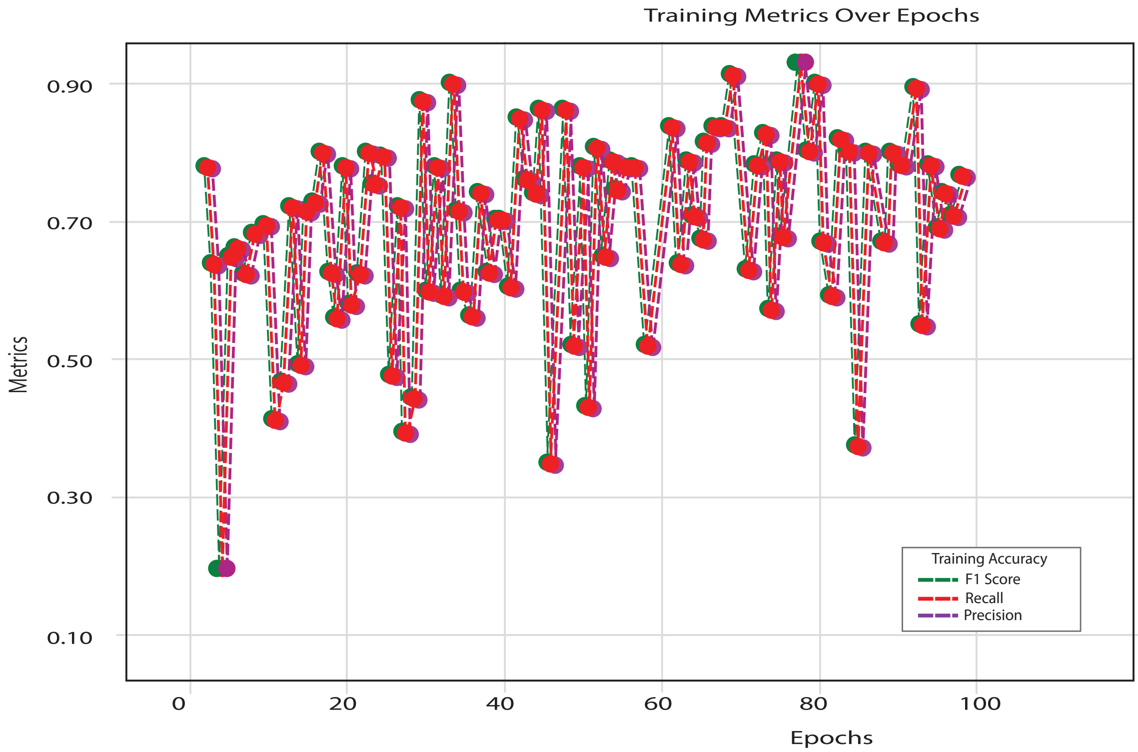

Figure 5.4: Evaluation Matrix for Expression Model

gesting that while the model becomes more accurate, it might still misidentify some cases as positive. The alignment of precision trends towards the end suggests that adjustments or improvements in data handling could mitigate these issues.

5.2.5 Real-Time Meltdown Detection and Emotion Analysis

This image illustrates our system's real-time detection of a meltdown, providing a comprehensive example of its capabilities. The system continuously monitors the patient's facial expressions and emotional states. In this instance, a meltdown is detected when the patient's hands move to their face or head, combined with specific emotional cues such as anger, fear, and sadness. The system tracks the patient's gaze, noted as "Center," to provide context for the emotional state and environment.

Upon detecting a meltdown, the system triggers an alert and provides a detailed breakdown of the patient's emotional state, enabling caregivers and medical professionals to understand the immediate context and potential triggers. This advanced monitoring and analysis allow for timely and accurate interventions. After a meltdown is detected, the system suggests personalized rehabilitation strategies tailored to the patient's specific needs and responses. These rehabilitation strategies may include therapeutic interventions, environmental modifications, or other supportive

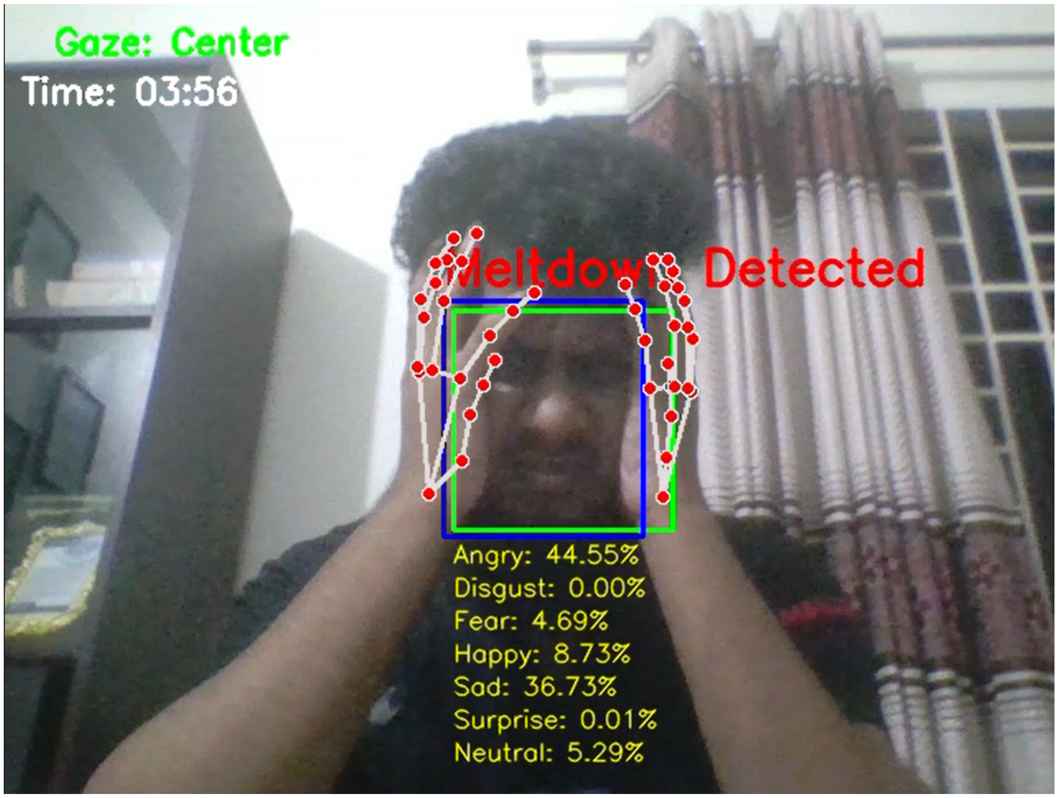

Figure 5.5: Our Model Result

measures aimed at mitigating or preventing future meltdowns. This functionality highlights the system's potential to significantly improve the quality of care and support for individuals with autism by providing continuous, real-time monitoring and personalized management plans.

5.3 Discussion

The results from the expression model highlight its potential in effectively recognizing and classifying facial expressions, which is crucial for the emotion detection module. The steady improvement in training accuracy suggests that the model benefits from the diversity and richness of the training data. However, the significant fluctuations in early epochs indicate the need for further optimization, possibly through hyperparameter tuning or regularization techniques to smooth the learning curve.

The autism detection model's performance, on the other hand, indicates room for improvement. The early plateau in accuracy suggests that the current feature set or the model complexity might be insufficient for capturing the nuanced patterns necessary for reliable autism detection. Enhancing the dataset with more varied and representative samples, or employing more complex models such as deep neural networks, could potentially yield better results.

The evaluation metrics provide a balanced view of the SVM model's strengths and weaknesses. While the consistency in F1 Score and Recall is promising, the variability in Precision highlights the need for further refinement to reduce false positives. This could involve more sophisticated feature engineering or incorporating additional data sources to improve the model's discriminative power.

In conclusion, while the expression model shows strong potential for emotion detection, the autism detection model requires further development to achieve comparable performance. Future work should focus on refining the feature set, exploring more advanced modeling techniques, and augmenting the dataset to improve overall accuracy and robustness.

CHAPTER VI

Conclusion and Future Work

6.1 Conclusion

In conclusion, our proposed autism detection and management system demonstrates significant potential in enhancing the early detection, monitoring, and personalized care for individuals with autism. By integrating advanced machine learning models and computer vision techniques, our system effectively recognizes and analyzes facial expressions, hand gestures, and emotional states in real-time. The expression model, based on a Convolutional Neural Network (CNN), shows strong accuracy improvement, indicating its efficacy in emotion detection. However, the Support Vector Machine (SVM) model for autism detection, while stable, suggests room for improvement in terms of feature selection and model complexity.

The system's ability to detect meltdowns through real-time monitoring of hand movements and emotional cues, and to subsequently suggest tailored rehabilitation strategies, represents a comprehensive approach to managing autism. The continuous feedback loop ensures that the system adapts to the evolving needs of the patient, thereby improving its effectiveness over time.

Overall, while the expression model's performance is promising, further refinement is needed for the autism detection component. Future work should focus on enhancing the dataset, exploring more sophisticated modeling techniques, and improving feature engineering to achieve higher accuracy and reliability. The integration of these improvements will help in providing more accurate and personalized care, ultimately contributing to better management and support for individuals with autism.

6.2 Future Work

Refining the SVM and CNN models to improve accuracy and robustness is a primary focus. This can be achieved through advanced techniques like deep learning and ensemble methods. Incorporating features such as voice tone analysis, eye-tracking, and body posture will provide a comprehensive understanding of patient behavior and emotions. Expanding the dataset to include diverse behaviors, age groups, and cultural backgrounds is essential. Real-time data collection and annotation, involving caregivers and professionals, will enhance adaptability. Optimizing algorithms for faster processing will minimize latency and improve user experience. Implementing edge computing solutions can process data locally, reducing internet dependency and improving response times. Designing intuitive interfaces for caregivers and professionals is crucial, and developing a mobile application will allow for on-the-go monitoring and intervention.

To enhance the system's effectiveness, implement adaptive learning techniques that tailor interventions based on individual progress and feedback. Customizing therapeutic videos and intervention strategies based on patient preferences will also be beneficial. Seamless data sharing and comprehensive care management can be achieved by integrating the system with existing healthcare platforms and electronic health records (EHR). Connecting with IoT devices, like smart home systems, will create a more supportive environment for patients. Improving behavior and emotion detection can be accomplished by integrating data from audio, visual, and physiological sensors. Context-aware algorithms that consider environmental factors will further enhance the system. Utilizing advanced machine learning techniques, such as reinforcement learning and transfer learning, can boost capabilities. Collaborations with behavioral scientists and conducting Human-Computer Interaction (HCI) studies will refine detection algorithms and improve system interaction. Focusing on these areas will ensure continuous improvement, providing better support for individuals with autism and their caregivers.

References

[1] B. H. Lo, F. Klopper, E. H. Barnes, and K. J. Williams, "Autism spectrum disorder," *Journal of Paediatrics and Child Health*, vol. 54, 2018.

[2] K. S. Paul, "Surviving meltdowns that cannot be prevented: Review of gaps in managing uncertainty and addressing existential vulnerabilities," *Journal of Risk and Financial Management*, 2022.

[3] N. Akhtar and M. Feeney, "Predictive analytics using a machine learning model to recommend the most suitable intervention technology for autism related deficits," in *2020 31st Irish Signals and Systems Conference (ISSC)*, pp. 1–6, IEEE, 2020.

[4] O. Altay and M. Ulas, "Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and k-nearest neighbor in children," in *2018 6th international symposium on digital forensic and security (ISDFS)*, pp. 1–4, IEEE, 2018.

[5] S. R. Dutta, S. Giri, S. Datta, and M. Roy, "A machine learning-based method for autism diagnosis assistance in children," in *2017 International Conference on Information Technology (ICIT)*, pp. 36–41, IEEE, 2017.

[6] F. C. Tamilarasi and J. Shanmugam, "Convolutional neural network based autism classification," in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pp. 1208–1212, IEEE, 2020.

[7] Y. Zhou, F. Yu, and T. Duong, "Multiparametric mri characterization and prediction in autism spectrum disorder using graph theory and machine learning," *PloS one*, vol. 9, no. 6, p. e90405, 2014.

[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] S. Ludyga, M. Gerber, F. Bruggisser, R. Leuenberger, M. Brotzmann, S. Trescher, M. Förster, L. Zou, E. Herbrecht, and M. Hanke, "A randomized cross-over trial investigating the neurocognitive effects of acute exercise on face recognition in children with autism spectrum disorder," *Autism Research*, 2023.

[10] S. K. Jarraya, M. Masmoudi, and M. Hammami, "A comparative study of autistic children emotion recognition based on spatio-temporal and deep analysis of facial expressions features during a meltdown crisis," *Multimedia Tools and Applications*, vol. 80, pp. 83–125, 2021.

[11] S. A. A. Shah, M. Bennamoun, and M. K. Molton, "Machine learning approaches for prediction of facial rejuvenation using real and synthetic data," *IEEE Access*, vol. 7, pp. 23779–23787, 2019.

[12] S. Rane, E. Jolly, A. Park, H. Jang, and C. Craddock, "Developing predictive imaging biomarkers using whole-brain classifiers: Application to the abide i dataset," *Research Ideas and Outcomes*, vol. 3, p. e12733, 2017.

[13] K. Hyde, A.-J. Griffiths, C. Giannantonio, A. Hurley-Hanson, and E. Linstead, "Predicting employer recruitment of individuals with autism spectrum disorders with decision trees," in *2018 17th IEEE international conference on machine learning and applications (ICMLA)*, pp. 1366–1370, IEEE, 2018.

[14] L. Gong, Y. Yan, J. Xie, H. Liu, and X. Sun, "Prediction of autism susceptibility genes based on association rules," *Journal of neuroscience research*, vol. 90, no. 6, pp. 1119–1125, 2012.

[15] K.-H. Cho, J.-B. Park, and A. Kang, "Metaverse for exercise rehabilitation: Possibilities and limitations," *International Journal of Environmental Research and Public Health*, vol. 20, no. 8, p. 5483, 2023.

[16] I. A. Ahmed, E. M. Senan, T. H. Rassem, M. A. Ali, H. S. A. Shatnawi, S. M. Alwazer, and M. Alshahrani, "Eye tracking-based diagnosis and early detection of autism spectrum disorder using machine learning and deep learning techniques," *Electronics*, vol. 11, no. 4, p. 530, 2022.

[17] Z. Zhao, H. Tang, X. Zhang, X. Qu, X. Hu, and J. Lu, "Classification of children with autism and typical development using eye-tracking data from face-to-face conversations: Machine learning model development and performance evaluation," *Journal of Medical Internet Research*, vol. 23, no. 8, p. e29328, 2021.

[18] S. Aminoleslami, K. Maghooli, N. Sammaknejad, S. Haghipour, and V. Sadeghi-Firoozabadi, "Classification of autistic and normal children using analysis of eye-tracking data from computer games," *Signal, Image and Video Processing*, vol. 17, no. 8, pp. 4357–4365, 2023.

[19] A. Capodieci, M. Romano, E. Castro, M. C. Di Lieto, S. Bonetti, S. Spoglianti, and C. Pecini, "Executive functions and rapid automatized naming: A new telerehabilitation approach in children with language and learning disorders," *Children*, vol. 9, no. 6, p. 822, 2022.

[20] A. Pahlevanian, N. Alirezaloo, S. Naghel, F. Alidadi, V. Nejati, and M. Kianbakht, "Neurofeedback associated with neurocognitive-rehabilitation training on children with attention-deficit/hyperactivity disorder (adhd)," *International Journal of Mental Health and Addiction*, vol. 15, pp. 100–109, 2017.

[21] L. F. Barrett, B. Mesquita, K. N. Ochsner, and J. J. Gross, "The experience of emotion," *Annu. Rev. Psychol.*, vol. 58, pp. 373–403, 2007.

[22] D. Leony, P. J. Muñoz-Merino, A. Pardo, and C. D. Kloos, "Provision of awareness of learners' emotions through visualizations in a computer interaction-based environment," *Expert Systems with Applications*, vol. 40, no. 13, pp. 5093–5100, 2013.

[23] M. Batty and M. J. Taylor, "Early processing of the six basic facial emotional expressions," *Cognitive brain research*, vol. 17, no. 3, pp. 613–620, 2003.

[24] M. I. U. Haque and D. Valles, "A facial expression recognition approach using dcnn for autistic children to identify emotions," in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pp. 546–551, IEEE, 2018.

[25] S. Karim, N. Akter, and M. J. Patwary, "Predicting autism spectrum disorder (asd) meltdown using fuzzy semi-supervised learning with nnrw," in *2022 International Conference on Innovations in Science, Engineering and Technology (ICISET)*, pp. 367–372, IEEE, 2022.