FIRAT UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
TÜRKİYE

AUTISM DETECTION FROM FACIAL IMAGES USING DEEP LEARNING METHODS

Abdulazeez Mousa ALMAHMOOD

Master's Thesis

DEPARTMENT OF SOFTWARE ENGINEERING
Program of Graduate School of Natural and Applied sciences

JANUARY 2022

FIRAT UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
TÜRKİYE

Department of Software Engineering
Graduate School of Natural and Applied sciences

Master's Thesis

AUTISM DETECTION FROM FACIAL IMAGES USING DEEP LEARNING METHODS

Author
Abdulazeez Mousa ALMAHMOOD

Supervisor
Assoc. Dr. Fatih ÖZYURT

JANUARY 2022
ELAZIG

FIRAT UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
TÜRKİYE

Department of Software Engineering
Graduate School of Natural and Applied sciences

Master's Thesis

Title:	Autism detection from facial Images using deep learning methods
Author:	Abdulazeez Mousa ALMAHMOOD
Submission Date:	13 December 2021
Defense Date:	20 January 2022

THESIS APPROVAL

This thesis, which was prepared according to the thesis writing rules of the Graduate School of Natural and Applied Sciences, Fırat University, was evaluated by the committee members who have signed the following signatures and was unanimously approved after the defense exam made open to the academic audience.

Signature

Supervisor:	Assoc. Dr. Fatih ÖZYURT Firat University, Faculty of Engineering	Approved
Chair:	Dr. Öğr. Üyesi Soner KIZILOLUK Malatya Turgut Özal University, Faculty of Engineering and Natural Science	Approved
Member:	Doç. Dr. Muhammed TALO Firat University, Faculty of Engineering	Approved

This thesis was approved by the Administrative Board of the Graduate School on

.... / .... / 20 ....

Signature

Prof. Dr. Kürşat Esat ALYAMAÇ
Director of the Graduate School

DECLARATION

I hereby declare that I wrote this Master's Thesis titled “Autism detection from facial Images using deep learning methods ” in consistent with the thesis writing guide of the Graduate School of Natural and Applied Sciences, Fırat University. I also declare that all information in it is correct, that I acted according to scientific ethics in producing and presenting the findings, cited all the references I used, express all institutions or organizations or persons who supported the thesis financially. I have never used the data and information I provide here in order to get a degree in any way.

20 January 2022

Abdulazeez Mousa ALMAHMOOD

PREFACE

I would like to dedicate this dissertation to my family. A special dedication to my mother and my father, who encouraged me to further my study and career, and who also, taught me to never stop setting and reaching newer and bigger goals. I dedicated to my dear supervisor Assoc. Dr. FATIH ÖZYURT

Abdulazeez Mousa ALMAHMOOD
ELAZIG, 2022

TABLE OF CONTENTS

Preface. iv
Table of Contents v
Abstract vi
Özet.... vii
List of Figures viii
Lıst of Tables ix
Abbreviations x
1. INTRODUCTION 1
Literature Review 5
Study purpose 7
Outline of the thesis 8
2. THEORETICAL BACKGROUND 9
Autism.. 9
Causes of Autism 10
Autism diagnosis 10
Deep Learning 10
Classification 12
Deep Learning in medical images field 13
Convolution Neural Network 13
Convolutional layer 14
Activation Layer 15
Pooling layer 16
Fully connected layer 16
Dropout 16
Batch Normalization 17
CNN architectures 18
Inception V3 18
DenseNet121 19
ResNet50 20
Xception 20
MobileNet 21
Support Vector Machines 22
3. MATERIAL AND METHOD 23
Dataset. 23
Data Augmentation 24
Transfer learning 25
Evaluation metrics 27
Hardware and software configuration 30
4. RESULT AND DISCUSSION 32
Experimental results 32
Discussion 42
CONCLUSION 43
REFERENCES 44
Currıculum Vıtae

ABSTRACT

Autism detection from facial Images using deep learning methods

Abdulazeez Mousa ALMAHMOOD

Master's Thesis

FIRAT UNIVERSITY
Graduate School of Natural and Applied Sciences

Department of Software Engineering
Graduate School of Natural and Applied sciences
January 2022, Page: x + 46

Autism spectrum disorder (ASD) refers to a collection of behavioral and developmental issues and difficulties. The cognitive, communication, and play skills of a child with autism spectrum disorder are all affected. The description "spectrum" in autism spectrum disorder refers to the fact that each child is special and has their own set of characteristics different from other children. These come together to give him a unique social bond as well as his own understanding of his own actions. Medical image classification is a significant research field that is gaining traction among researchers and clinicians alike to detect and diagnose diseases. It addresses the issue of medical diagnosis, experiment purposes and analysis in the field of medicine. Several deep learning-based medical imaging applications have been proposed and developed to understand and learn about how diseases develop in patients, to help doctors in early diagnosis of pathology. Not only is achieving good accuracy in classifying medical images the main purpose alone. We used pre-trained CNN and transfer learning in this study. These CNNs pre-trained architectures are used to train the network and to classify medical images. The experimental results of this study show that the model based on transfer learning (Support Vector Machine + MobileNet) can detect Autism Spectrum Disorder at the best rate with 98.83% accuracy. The architectures that were tested in this study are ready to be tested with additional data and can be used to prescreen individuals with ASD. The use of deep learning methods for feature selection and classification in this study could greatly support future autism studies.

Keywords: Autism Spectrum Disorder (ASD); Transfer Learning; Deep learning; pre-trained model; CNN; Support Vector Machine (SVM).

ÖZET

Derin öğrenme yöntemlerini kullanarak yüz görüntülerinden otizm tespiti

Abdulazeez Mousa ALMAHMOOD

Yüksek Lisans Tezi

FIRAT ÜNİVERSİTESİ
Fen Bilimleri Enstitüsü

Yazılım Mühendisliği Anabilim Dalı
Fen Bilimleri Enstitüsü
Ocak 2022, Sayfa: x + 46

Otizm spektrum bozukluğu (OSB), davranışsal ve gelişimsel sorunların ve zorlukların bir koleksiyonunu ifade eder. Otizm spektrum bozukluğu olan bir çocuğun bilişsel, iletişim ve oyun becerilerinin tümü etkilenir. Otizm spektrum bozukluğundaki "spektrum" tanımı, her çocuğun özel olduğu ve diğer çocuklardan farklı olarak kendine özgü özelliklere sahip olduğu gerçeğini ifade eder. Ona kendi eylemlerine ilişkin kendi anlayışının yanı sıra benzersiz bir sosyal bağ için bir araya gelir. Tıbbi görüntü sınıflandırması, hastalıkları tespit etmek ve teşhis etmek için hem araştırmacılar hem de klinisyenler arasında ilgi çeken önemli bir araştırma alanıdır. Tıp alanında tıbbi teşhis, deney amaçları ve analiz konularını ele alır. Hastalarda hastalıkların nasıl geliştiğini anlamak ve öğrenmek, doktorlara patolojinin erken teşhisinde yardımcı olmak için bir dizi derin öğrenme tabanlı tıbbi görüntüleme uygulaması önerilmiş ve geliştirilmiştir. Tıbbi görüntülerin sınıflandırılmasında, iyi bir doğruluk oranı elde etmek ana amaç değildir. Bu çalışmada, önceden eğitilmiş evrişimsel sinir ağlarını ve transfer öğrenmeyi kullandık. Önceden eğitilmiş evrişimsel sinir mimarileri, ağı eğitmek ve tıbbi görüntüleri sınıflandırmak için kullanılmaktadır. Bu çalışmanın deney sonuçları, önerilen modelin, transfer öğrenmeye dayalı Destek Vektör Makinesi (DVM) ile MobileNet modeli kullanılarak elde edilen en iyi yüzde 98,83 doğrulukla OSB tespit edebildiğini göstermektedir. Bu çalışmada test edilen mimariler, ek verilerle test edilmeye hazırdır ve OSB'li bireyleri ön tanı için kullanılabilir. Bu çalışmada öznitelik seçimi ve sınıflandırma için derin öğrenme yöntemlerinin kullanılması, gelecekteki otizm çalışmalarını büyük ölçüde destekleyebilir.

Anahtar Kelimeler: Otizm spektrum bozukluğu; Transfer Öğrenimi; Derin Öğrenme; Önceden Eğitilmiş Model; Evrişimsel Sinir Ağları (ESA); Destek Vektör Makinesi (DVM).

LIST OF FIGURES

Page

Figure 2.1 Deep Learning ... 12

Figure 2.2. CNN convolution operation ... 15

Figure 2.3. Dropout technique in Deep Learning ... 17

Figure 3.1. Autistic and Non-Autistic samples ... 23

Figure 3.2. Procedure of rotating for image at various angles (Rotation Transformation) ... 25

Figure 3.3. Pre-trained models are used in the transfer learning procedure. ... 27

Figure 4.1. The Confusion Matrix of MobileNet with SVM classifier. ... 33

Figure 4.2. The Learning Curve of MobileNet with SVM classifier. ... 33

Figure 4.3. The Loss Function of MobileNet with SVM classifier. ... 34

Figure 4.4. The Confusion Matrix of MobileNet model. ... 34

Figure 4.5. The Learning Curve of MobileNet model. ... 35

Figure 4.6. The Loss Function of MobileNet model. ... 35

Figure 4.7. The Confusion Matrix of Densenet121 model. ... 36

Figure 4.8. The Learning Curve of Densenet121 model. ... 36

Figure 4.9. The Loss Function of Densenet121 model ... 37

Figure 4.10. The Confusion Matrix of Inception V3 model. ... 37

Figure 4.11. The Learning Curve of Inception V3 model. ... 38

Figure 4.12. The Loss Function of Inception V3 model. ... 38

Figure 4.13. The Confusion Matrix of Xception model. ... 39

Figure 4.14. The Learning Curve of Xception model. ... 39

Figure 4.15. The Loss Function of Xception model. ... 40

Figure 4.16. The Confusion Matrix of ResNet-50 model ... 40

Figure 4.17. The Learning Curve of ResNet-50 model. ... 41

Figure 4.18. The Loss Function of ResNet-50 model. ... 41

LIST OF TABLES

Page

Table 4.1. Hardware identifications 30

Table 4.2. CNN network’s training option hyper-parameters and values 31

Table 4.3 Pre-trained networks’ results 32

ABBREVIATIONS

DL	: Deep Learning
TL	: Transfer Learning
AI	: Artificial Intelligence
ML	: Machine Learning
NN	: Neural Network
FC	: Fully Connected
SVM	: Support Vector Machine
CNN	: Convolutional Neural Network
ASD	: Autism Spectrum Disorder

1. INTRODUCTION

Every person on the planet has their own existence and being, which they used to contribute to numerous social and practical duties. Every community has a specific category of people who as a cause of differences in their health from other normal people require a special adaptation to the environment in which they live. Due to the prevalence of autism in a huge number of children around the world, it has unquestionably become a requirement of existence. Autism is one of the most perplexing disorders and impairments, and its importance is owing to the ambiguity of this term for many people of various social and cultural groups [1]. There are several forms of autism, including the presence of permanent developmental disorders in the child. The presence of unspecified growth disorders. The presence of regressive childhood disorders. Aspects of communication and social interaction in autistic patients The child shows some symptoms that indicate autism in the field of communication and social interaction, such as: The child does not respond when his or her name is called. He refuses to be hugged or held, as he often prefers to be alone. There is a delay in speaking with a loss of pronouncing words correctly. The inability to start a conversation with others and continue this conversation well. The child speaks in an abnormal tone of rhythm and has a monotonous voice. The child does not express his or her feelings clearly. The child reacts in an aggressive or disruptive manner in a disruptive and inappropriate manner. A kid with autism exhibits a variety of behaviors. The following are some of the behavior patterns that can be seen in a child with autism: He or she bites himself or hits his or her or her head against the wall, among other things that may be harmful to him or her. This kid has some imbalanced motor patterns, such as walking on his toes or other similar behaviors. He can be enthralled by little details like witnessing the wheels rotate rather than the big image. Some senses, including hearing, touch, light, and others, are hypersensitive in this kid. may choose unusual cuisine and refuse to consume familiar foods. Autism interferes with a child's normal brain development in the areas of social interaction and communication. Nonverbal communication, social engagement, and leisure activities are all common issues for children and people with autism. Autism makes it difficult for people to communicate with one another and relate to the outside world. People with this disorder may engage in abnormally repetitive behavior, such as flapping their hands or shaking their bodies frequently, and they may have unusual responses when dealing with people or abnormally associate with certain things, such as a child playing with a specific car repeatedly and abnormally, without attempting to switch to another car or toy, and with resistance to change. The youngster may act aggressively against others or even himself in some instances [2].

A child with autism does not learn on his or her own, but rather through indoctrination, and he or she is curiously inventive, routine, and does not express himself or herself with words, instead of relying on his or her forceful approach, which appears juvenile or childish fury. Autism develops as a result of a digestive issue that affects the nervous system via the bloodstream, disrupting brain function. As a result, people with autism are unable to develop information in the same way that normal people do, and there is no medication cure for this disorder. And while many autistic children are unable to adapt to life, they can do so through training, they will not be able to understand life normally or coexist with the society in which they live without it. Their tendency to independence, which is one of the most noteworthy manifestations associated with puberty in people with autism, requires assistance because it is not as strong as it is in typical people. Their perception isn't as sharp as others; therefore, it takes a lot of work to do it right. Because of the social hurdles he confronts in his life, which prohibit him from acting as he likes, an autistic youngster is under a considerable degree of psychological stress. When an autistic child reaches adolescence, he or she will go through changes that are different from those he or she went through as a child [3]. Since the family finds difficulties and does not have the ability to control him and his actions because he is no longer a child, the family of a child with autism faces a number of problems, including dealing with him and communicating his condition to others and the community. And the family's difficulty in preventing an autistic son or daughter from engaging in socially inappropriate behaviors. direct cause of autism has yet to be determined through scientific researches. Although most studies suggest that a genetic element directly influences the development of this condition because identical twins (from the same egg) are more common than other twins (from two different eggs), and identical twins are known to have the same genetic composition. Recent radiography, such as magnetic frequency imaging, revealed the presence of several aberrant signals in the structure of the brain, with distinct changes in the cerebellum, including differences in the brain's size and the number of "Perkinje cells". Because a genetic element is the most likely cause of autism.

Treatment and Intervention Services for Autism Spectrum Disorder [4]. there is currently no cure for autism spectrum disorder (ASD). However, numerous therapies for use with youngsters have been created and evaluated. These interventions may improve a child's ability to function and engage in society by reducing symptoms, improving cognitive ability, and daily living skills. Children with autism have unique strengths and limitations in social communication, behavior, and cognitive capacity due to the variances in how autism spectrum disorder impacts each person. As a result, treatment regimens are typically interdisciplinary, may involve parent-mediated interventions, and are tailored to the child's specific requirements. Behavioral intervention tactics have emphasized the development of social communication skills, especially at a young age when these skills are naturally acquired, as well as the reduction of restricting interests and recurring behavioral issues. Occupational therapy and speech therapy, as well as social skills training and medication, may be beneficial to some children. The optimal treatment or intervention depends on the age, strengths, challenges, and differences of the individual. It's also crucial to realize that children with autism, like children without autism, can get sick. As a result, medical and dental exams should be included in the child's treatment plan. It's often difficult to tell whether a child's conduct is caused by autism spectrum disorder or by anything else. Head banging, for example, could be a symptom of autism spectrum condition or a sign that a youngster is suffering from headaches or earaches. A comprehensive physical examination is essential in these situations. Monitoring a child's healthy growth entails not only paying attention to autism symptoms but also to the child's physical and emotional well-being. The best interventions for older children and individuals with autism spectrum disorder are currently unknown. There has been some research on the social skills of older children, but there is not enough evidence to support it. More study is needed to assess programs aimed at improving adult outcomes. Furthermore, supports are necessary to assist people with autism with finishing their school or job training, finding jobs, securing housing and transportation, maintaining their health, improving everyday functioning, and participating as fully as possible in their communities [4], [5].

Diagnosing autism: this is one of the most difficult and complex issues, as the number of people who are scientifically prepared to diagnose autism is decreasing, which can lead to a diagnosis error or to ignoring autism in the early stages of a child's life, which makes intervening in times of crisis difficult. A child's behavior and communication abilities must be carefully observed and compared to typical levels of growth and development before a diagnosis can be made. However, many autistic behaviors are also observed in other illnesses, making identification even more challenging. As a result, in ideal circumstances, the child's condition should be assessed by a multidisciplinary team that includes a neurologist, a psychologist or psychiatrist, a developmental pediatrician, a language therapist and speech pathologist, an occupational therapist, an educational specialist, and other autism experts.

Although most studies, or a large proportion of studies on autism, have reached conclusions that the cause of autism is closer to genetic causes, or in other words, the disease is considered genetic. Therefore, when the child's behavioral qualities and facial features are examined, the highest rates of accurate diagnosis occur. Patients share a consistent pattern of unique facial malformations, allowing experts to diagnose the disease just on a single photograph of the kid. As a result, using computer technology to interpret the children's image model can considerably enhance the speed and accuracy of reading the picture, which will become interesting research direction in autism detection and diagnosis [6].

Deep learning has quickly become one of the most popular fields of research in computer science and applications. As a result of advancements in deep learning. Deep learning advances have contributed to the growth of neural network algorithms that can now compete with humans in vision tasks like image detection, classification, and segmentation. Imaging analysis in the medical image sector has also advanced tremendously as a result of the translation of these approaches into clinical science. This has developed a need in the medical image field for a better understanding of the techniques behind neural networks and deep learning, as close collaboration between computer scientists and medical image field researchers, as well as realistic expectations about the capabilities (and limitations) of the current state-of-the-art, is critical for successful implementation of deep learning techniques. In the field of medical imaging, deep learning techniques such as CNN are widely employed. However, assumptions made when defining the elements frequently limit them. Sensitivity suffers as a result of this. Deep learning algorithms, on the other hand, may be an appropriate answer because they can discover characteristics from raw image data. The absence of labelled medical image data is one of the problems in deploying these algorithms. Although this is a limitation that applies to all deep learning applications, the restriction on medical picture data is especially important owing to patient confidentiality [7].

In this study, the use of a standard dataset was proposed to test the competitors more accurately in order to identify the children as Autistic or Non-Autistic. In our study, we used a dataset of medical images of children's faces.

Autism patients require specialized medical care as well as a variety of integrated services such as health promotion and rehabilitation. Collaboration with other sectors, particularly education, rehabilitation and social care centers, is crucial to the health sector. Therefore, the importance of early childhood treatment in the development and well-being of children with autism spectrum disorders lies in strengthening its treatment methods and how to adapt and engage with the community, family and friends around it. Baby growth should be monitored as part of standard maternal and child health care. Important information, advice, and practical help for children and adults with autism, as well as their care, should be supplied in response to their changing unique needs and preferences. Interventions for people with autism should be created and conducted with the assistance of their specialists for improved facilitation, inclusion, and support, and social and community measures should be included in the treatment.

To analyze and detect ASD at an early stage, we suggested a transfer learning-based autism facial detection framework in this study. Therefore, first we employed deep learning pre-trained models for classification on our dataset that contains medical images. And then in the second time we combined machine learning classifier with deep learning pre-trained models to get better results. We used pre-trained models for feature extraction and took it as input for a machine learning classifier. Miss rate became lower after the combination of deep learning pre-trained models with machine learning classifier compared to the implementation of the other deep learning pre-trained models. This type of model can quickly provide a diagnostic test for probable autism features in the facial image. In summary, our research is essentially an image binary classification based on transfer learning and autism can be detected from such a child's input images.

Literature Review

In the detection, diagnosis, and treatment of various diseases, medical imaging technology is critical. Deep learning technology is predicted to aid researchers and physicians in enhancing the accuracy of imaging diagnosis and treatment while also lowering the imbalance of medical resources due to the volatility of human expert expertise. Deep learning has been used by many academics in the field of medical image analysis.

M. Beary et al. [8] Proposed diagnosis of autism in children using facial analysis and deep learning. In this paper, the authors proposed a deep learning model to classify children if they are autistic or non-autistic (healthy). Autistic patients in general have distinct facial abnormalities that are part of the disease discovery that allows researchers to know if a person has autism or not. Therefore, the researchers in this study used the MobileNet deep learning model with two dense layers to extract the features of the face and classify the people according to the image. In this study, the model was trained and tested on 3014 images (90% for training, 10% for testing). The results of this study were 94.6% accuracy in classifying.

M. S. Mythili et al. [9] have written a study on autism spectrum disorders using classification techniques. In this study, the authors focused on extracting data from methodologies to study the performance of students with autism. By mining data with the tasks, it provides that can be used to study the performance of students with autism. In this article the authors use algorithms for data mining by task classification at the level of students. The authors used Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Fuzzy logic as machine learning methods in this paper. The algorithms are extremely helpful in dealing with the predicted level of autistic students.

N. Ali et al. [10] have written autism spectrum disorder classification on electroencephalogram signal using deep learning algorithm in this paper, the authors use deep learning algorithms via electroencephalogram (EEG) to detect the different patterns between normal children and children with autism. The process of classifying normal children and children with autism by using deep learning algorithms and using the Multilayer Perceptron (MLP) model in this article is by means of a database that contains brain signals for pattern recognition, using the Multilayer Perceptron (MLP) model to extract the features for the classification process.

N. Rad et al. [11] have written deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders. In this paper the authors proposed deep learning application to aid automatic Stereotypical Motor movements (SMM) detection with multi-axis Inertial Measurement Units (IMUs). In the study from raw data, CNN was used to learn a discriminative feature space, this method is used to model the temporal patterns in a series of multi-axis signals, combine the long short-term memory (LSTM) with CNN architectures. Furthermore, it was demonstrated how the CNN can be used with a transfer learning technique to improve the detection and analysis rate on longitudinal data. From the results of this paper they show that handcrafted features are outperformed by feature learning, the detection rate improves when using LSTM to learn the temporal dynamic of signals, particularly when the training data is distorted. Detectors with an ensemble of LSTMs are more accurate and stable. In longitudinal settings, parameter transfer learning is beneficial. These results represent a significant step forward in detecting SMM in real-time scenarios.

T. Akter et al. [12] proposed improved transfer-learning-based facial recognition framework to detect autistic children at an early stage. An effective transfer-learning based on autism face detection framework is suggested in this experiment to more accurately identify children with autism spectrum disorder (ASD) in the early years. With using dataset that contains children's faces images with ASD from Kaggle. In this study, many deep learning and machine learning classifiers were implemented by applying pre-trained models based on transfer learning. That noted, the value of both the falling and error rate is the lowest and the highest accuracy of the improved MobileNet-V1 model when compared to other pre-trained models and classifiers. Furthermore, for using the k-means clustering technique, this classification model is used to determine the various autism spectrum disorder (ASD) groups studying solely autism image data.

S. Raj et al. [13] have proposed analysis and detection of autism spectrum disorder using machine learning techniques. In this paper various machine learning algorithms used for prediction and analysis autism spectrum disorder (ASD) with child, and adults. algorithms implemented in this research: Naïve Bias (NB), Support Vector Machine (SVM), Logistic Regression (LR), K-Nearest Neighbors, Neural Networks (NN), and Convolutional Neural Networks (CNN). In this study, three different datasets were used and applied by machine learning techniques. the results show that: better performing of Convolutional Neural Networks (CNN) based prediction models than others. With best accuracy of 99.53%, 98.30%, 96.88% of prediction for ASD.

B. Li et al. [14] applying machine learning to identify autistic adults using imitation: An exploratory study suggested. The purpose of this research was to look into the core issue of discriminative test conditions and kinematic characteristics. ASC individuals with a series of hand gestures are included in the dataset. Machine learning algorithms were used to extract 40 kinematic constraints from 8 imitation circumstances. This study shows that utilizing machine learning approaches to assess high-dimensional datasets and the definitive classification of autism is feasible for a small sample. RIPPER's sensitivities rates for the features Va (87.30%), CHI (80.95%), IG (80.95%), Correlation (84.13%), CFS (84.13%), and "no extraction of features" (80.00 percent) on the AQ-Adolescent dataset.

F. Thabtah et al. [15] proposed autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment. They have suggested a DSM-5-based ASD screening methodology based on Machine Learning Adaptation. In autism spectrum disorder (ASD) screening, a screening instrument was employed to achieve one or more aims. The researcher explored the advantages and disadvantages of autism spectrum disorder (ASD) Machine Learning classification in this work. The researcher attempted to emphasize the issues with existing ASD screening methods, as well as the consistency of such tools that use the DSM-IV rather than the DSM-5.

D. Stahl et al. [16] proposed novel machine learning methods for ERP analysis: a validation from research on infants at risk for autism. Using event-related potential data on eye gaze ability, researchers were able to distinguish between infant groups at high and low-risk for later autism diagnosis. Regularized discriminant functions analysis (DFA), linear discrimination analysis (LDA), and SVM were used to investigate computational approaches. SVM had a 64-percent accuracy, regularized DFA had a 61-percent accuracy, and LDA had a 56-percent accuracy. In order to improve predicted accuracy, they stressed the importance of eliminating irrelevant variables and identifying the most discriminative features prior to categorization.

J. Wang. et al. [17] proposed Multi-task diagnosis for autism spectrum disorders using multi-modality features: A multi-center study. For autism spectrum disorder(ASD) diagnosis, they suggest an innovative multi-modality multi-center classification (M3CC) technique. They categorize each imaging center's classification as a single task. They address the classification with all imaging centers at the same time by applying task-task and modality-modality regularizations. However, for highly accurate diagnosis, the best feature selection and discriminant function modeling can be done simultaneously. They also offer and examine the convergence of an effective iterative optimization solution to their specified problem. In comparison to existing methods, their exhaustive studies on the ABIDE database reveal that their suggested method can greatly increase the performance of ASD diagnosis.

The autism detection is given out in this study, proving the early detection of autism. Convolutional neural networks have been recommended for processing pre-trained networks on medical images. Some of the most popular pre-trained models, such as MobileNet, Densenet121, Inception V3, Xception, and Resnet50, were used alone at first, and then these pre-trained models were combined with a machine learning classifier called Support Vector Machine (SVM) and the results were compared with the results of first time to get the difference between using pre-trained models alone and the models with machine learning classifier to get the best results, and Transfer Learning was used to implement our thesis.

Study purpose

The goal of this thesis is to apply deep learning algorithms to determine whether or not images of children's faces in medical images include autism. Using deep learning models, in particular, ensures that a more accurate and robust structure is shown. For this procedure, classification algorithms will be studied, and the most famous deep learning architectures will be presented and then again one more time these deep learning algorithms will be combined with a machine learning classifier called Support Vector Machine (SVM) will be presented. The results will be presented in a comparable manner.

Outline of the thesis

The current research is presented as follow:

The theoretical background of the thesis is presented in Chapter two. In addition, brief description of the convolutional neural networks is presented, convolutional neural networks (CNN) architectures (MobileNet, DenseNet121, Inception V3, Xception, and ResNet50), and then the machine learning classifier called support vector machine (SVM) discussed, as well as explanations of deep learning and classification methods. It also contains details of the evaluation matrix and the deep learning architectures used in this study.

The approach employed in this thesis is presented in chapter three. In addition, this chapter included an explanation of deep learning approaches. The dataset and many evaluation methodologies utilized in the thesis will also be explained in Chapter Three, explaining in detail the software, hardware requirements that will be needed for the thesis.

The results of a study will be reported in chapter 4, along with a comparative to other existing methodologies. Furthermore, this chapter focuses on how well the network has tested the outcomes.

2. THEORETICAL BACKGROUND

Autism

Autism, also known as subjectivity, is one of the neurologic illnesses known as ASD and it causes a variety of symptoms, these symptoms which might appear generally in the first two years. Autism is a neurodevelopmental disorder indicated by lack of ability and difficulty interacting, communicating, and socially interacting with others, including verbal and nonverbal communication, difficulty understanding how others feel or think, anxiety and annoyance around social events and unfamiliar situations, and taking longer than usual to comprehend information. Bright lights irritate them, allowing them to feel stressed and uncomfortable. Interrupted speech, or the loss of the ability to do so after doing and thinking about the same things repeatedly. Autism alters data processing in the brain by altering the connections and organization of neurons. Autism symptoms arise in children throughout their first and second years of life, while the illness is sometimes diagnosed later. It is important to note that while there is no treatment for this illness, speech therapy, occupational therapy, and educational support, among other therapies, can aid this group of children and their families [18].

These four subgroups of autism spectrum disorder (ASD) classify autism:

- **Asperger's syndrome**

  Symptoms of Asperger's are frequently milder than those of other kinds of autism. It is characterized by trouble interacting in social situations, indifferent (caring) about others, difficulty reading others body language and expressions, lack of eye contact, repeated social activities, and inability to read social cues.

- **Autism spectrum disorder**

  Many problems occur during chats and discussions, and there are frequently occasional seizures, as well as repeating the same actions and behaviors, as well as difficulty sleeping, eating, and feeding habits.

- **Pervasive developmental disorder**

  This type of disorder has a number of chaotic speech patterns, as well as a high degree of accuracy in following a daily routine, taste, sight, smell, and voice.

- **Childhood disintegrative disorder**

  This disorder is sometimes characterized by a sharp end of the spectrum, weakness and trouble speaking, enuresis, and a person's motor abilities are underdeveloped.

Causes of Autism

Doctors There is no single element that has been identified as a conclusive cause of autism. Given the disease's intricacy, the breadth of autistic disorders, and the fact that two autistic states do not match, that is, between two autistic children, numerous elements are thought to be involved in the development of autism. Factors that are genetic. Several genes appear to play a role in the development of autism spectrum disease. Autism spectrum disorder can be connected to a genetic condition such as Rett syndrome or fragile X syndrome in some children. Other children's risk of acquiring autism spectrum disorder may be increased as a result of genetic variations (mutations). Other genes, on the other hand, may have an impact on brain development, the way brain cells connect or the intensity of symptoms. Some genetic mutations seem to be inherited, while others seem to occur on their own. Environmental effects Researchers are actively investigating whether viral infections, medicines, pregnancy difficulties, or air pollution plays a role in the development of autism spectrum disease [18], [19].

Autism diagnosis

The treating doctor monitors the child's growth and development on a regular basis to detect any developmental delays. If the child exhibits autism symptoms, he can seek treatment from an autism specialist who, in collaboration with a team of other professionals, will conduct an accurate assessment of the illness. Because autism has such a wide range of symptom severity, diagnosing can be a difficult and time-consuming procedure, especially because there is no specific medical examination that can detect the presence of autism. Diagnosis seeing the child's expert doctor is part of the formal examination of autism. Parental discussion of the child's social skills, language ability, and behavior, as well as how and to what extent these factors change and grow over time. Putting the child through a series of exams and tests in order to analyze his speech and language ability as well as look into certain psychological issues [20], [5]. Although the first signs of autism show at the age of eighteen months, the final diagnosis is often determined only when the child is two or three years old, when there is a developmental delay, a delay in language acquisition, or a deficit in interpersonal social relationships. Early diagnosis is critical because early intervention, particularly before the child reaches the age of three, is critical to achieve the greatest possible outcomes and prospects for improvement.

Deep Learning

Deep learning seeks to emulate the human mind, while it falls far short of its capabilities, allowing systems to classify data and generate extremely accurate predictions. Deep learning is a function of machine learning that is essentially a three-layer neural network. These neural networks aim to imitate the activity of the human brain by allowing it to "learn" from enormous amounts of data, although they fall far short of its capabilities. While a single-layer neural network may still produce approximation predictions, adding hidden layers can aid in accuracy optimization and refinement. Many artificial intelligence (AI) systems and services rely on deep learning to improve automation by executing analytical and physical activities without the need for human participation [21]. Everyday products and services (such as voice assistants, sound TV remotes, and card fraud detection) as well as upcoming innovations use deep learning technology (such as self-driving cars, diseases detection and so on). Deep learning neural networks, also known as artificial neural networks, use a combination of advanced inputs, weights, and bias to try to emulate the human mind. These pieces work together just to realize, classify, and characterize items in the dataset accurately. Deep neural networks are made up of numerous layers of nodes connected, each of which improves and refines the prediction or classification. Forward propagation refers to the progression of calculations via the network. The visual layers of a deep network are the input and output layers [7]. The input layer a set of input data, which is composed of values that are independent of one another and correspond to the input type. The output layer correlates to the relevant income circumstance, such as the expected weather condition (clear/rainy) or a patient's health state (healthy/sick).

Deep learning technology is a special technology that is very similar to the work and way of thinking of the human mind in understanding things through experiences. A deep neural network is a term used to describe artificial intelligence. The term "deep" refers to the hidden layers of a neural network. Deep learning models are taught using a large quantity of data and the data structure of neural networks, which learn features directly from the data without the need for user intervention. Neural networks, like our biological neural networks, are connected systems. These systems are built in such a way that they can adapt to changing circumstances. Once neural networks have determined the results of a certain object, they can tell whether or not it is the same object the next time around. Neural networks don't perceive items the same way we do; instead, they recognize objects based on their own set of characteristics [22].

Graphical represent relationship between deep learning, machine learning, and artificial intelligence shown below in Figure 2.1.

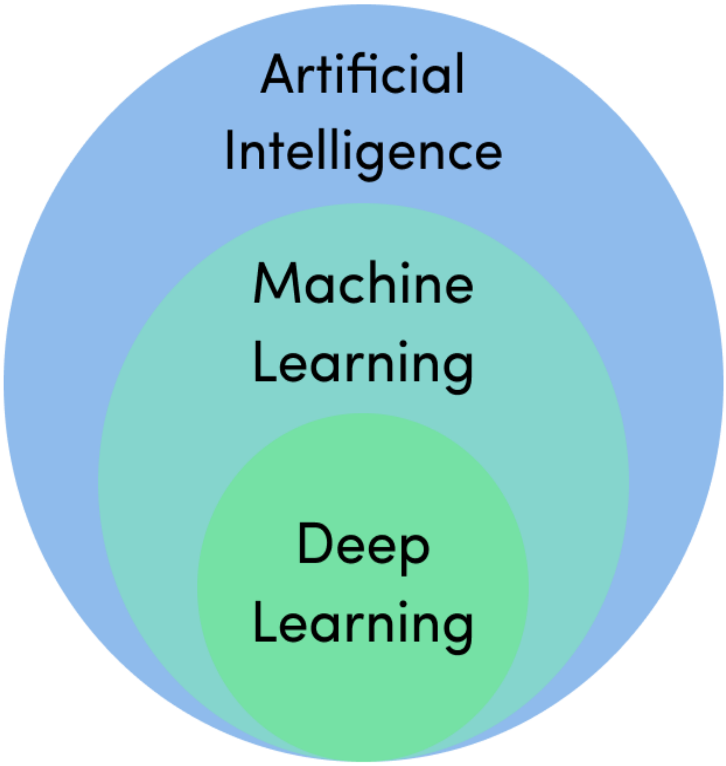

Figure 2.1 Deep Learning

Classification

Image classification is a subfield of computer vision in which a classifier examines an image and labels a tag to it from a set of predetermined labels or groups on which it has been trained. As a result, one of the first applications of computer vision that researchers considered was simulating a classification problem using neural networks [23].

Image Pixels are used to represent an image on a computer. In its view, a picture is nothing more than an array of matrices, the size of which is determined by the image resolution. The study of these mathematical representations with the use of algorithms is essentially image processing for the computer. The algorithms reduce the strain on the final classifier by breaking down the image into a set of its most noticeable features. These characteristics help the classifier figure out what the image depicts and which class it should be assigned to. The feature extraction method is the most important step in classifying a picture because it is the foundation for all subsequent phases. That data presented to the algorithm also plays a big role in classification, especially supervised classification. When compared to a terrible dataset with class-wise data imbalance and poor picture and annotated quality, a well-balanced classification dataset works wonders [23], [24].

Deep Learning in medical images field

Deep learning is composed of a number of layers that allow for various levels of abstraction and better data predictions. It is quickly becoming the most popular machine-learning technique in the overall imaging and computer vision fields. Accurate diagnosis in medical imaging analysis is dependent on both picture acquisition and image interpretation. In recent years, image acquisition has advanced significantly, with equipment gathering data at quicker rates and with higher resolution. Computer technology has just begun to improve the image interpretation process. This is soon to change thanks to deep learning and medical image computing. Convolutional neural networks (CNNs), for example, have previously proven to be an effective deep learning method for medical picture interpretation. Image analysis with deep learning. Even the most complex abstractions derived from medical imagery are automatically learned by CNNs [25].

In Deep learning CNNs medical image analysis is significant facilitators for enhancing diagnosis by making it easier to identify abnormalities that need to be treated and supporting the physician's workflow. Deep learning models are quicker, more accurate, and, most importantly, they are never-ending, unlike human doctors. As a result, they are far more effective than human doctors or radiologists. They can analyze thousands of images every day without showing any signs of exhaustion. Technology is linked to deep learning and machine learning. Machine learning is a subfield of artificial intelligence in particular. Similarly, it is linked to the development of AI algorithms that can modify (learn) on their own to achieve the intended result without the need for human interaction. Machine learning includes deep learning as a subfield. It is a more advanced technology in which algorithms are built and operate on multiple levels, each of which provides a distinct interpretation of the data. An artificial neural network is a collection of algorithms that replicates the neural connections seen in the human brain [25], [26].

In deep learning images are used as inputs to one or more images on a regular basis and then as output a single diagnostic value will be used to classify medical images (for example, whether there is a disease or the person is healthy). In the field of image classification [26], CNN is the current ideal. Imaging classification field, that can be separated into whole image classification and target classification, is one of the earliest uses of deep learning in the medical imaging area. The full image and its subcategory labels are typically used as input for the entire image categorization. It uses a deep learning method to perform end-to-end training to understand the data features of a certain category.

Convolution Neural Network

Deep learning techniques rely on neural networks, which are a subset of machine learning. They're made up of node levels, each of which has an input layer, one or many hidden layers, as well as an output layer. Each node is connected to the others and has a value and limit assigned to it. If a node's output exceeds a certain limit value, that node is activated, and data is sent to its next layer of the network. Otherwise, no information is being sent to the network's next layer. Neural networks come in a variety of shapes and sizes, and they're utilized for a variety of purposes and data kinds. Recurrent neural networks (RNN), for example, are frequently employed in speech recognition and natural language processing, but CNN are more frequently used in classification and computer vision applications. Detecting entities in images before CNNs required laborious, time-consuming feature extraction approaches. Convolutional neural networks, on the other hand, use concepts from linear algebra, notably matrix multiplication, to discover patterns inside an image, making them more scalable for image classification and object recognition applications. They can, however, be computationally intensive, necessitating the use of graphics processing units (GPUs) to train algorithms [21], [6]. The convolution layer shares all the weights by applying the convolution operation to the given input image. In reality, the convolutional layer's task is to detect and recognize features at various locations in the input image/feature map. The convolution neural network (CNN) becomes one of the most effective applications of deep learning is an area of diagnostic purposes in the field of medical image analysis.

Convolutional layer

The convolutional Convolution is a mathematical operator that builds a third function by combining two functions, f and g. To get a new output element, the convolution kernel and input elements are triggered.

Activation mapping (feature mapping): The filter (also known as neurons or nuclei) acts in the receptive field if the input is an array of values of 32 × 32 x 3 pixels (elements are called weights or parameters), The idea is that the filter depth must be the same as the depth of the input material (to allow for calculations), therefore the filter size is 5 x 5 x 3. The filter initially moves one pixel to the right in the upper left corner of the image. The value in the filter will be multiplied by the original pixel value in the image before each slice (also known as raster product arithmetic). To get a number, these items are added together. The result can be written as a 28 x 28 matrix because this filter can provide 28 × 28 outcomes.

The overall output becomes 28 x 28 x 2 when two filters are used instead of a single 5 x 5 x 3 filter. The greater the number of filters used, the better the spatial dimensions are kept. Each torsion core test has its own set of characteristics, in the Figure 2.2, convolution layer operation explained.

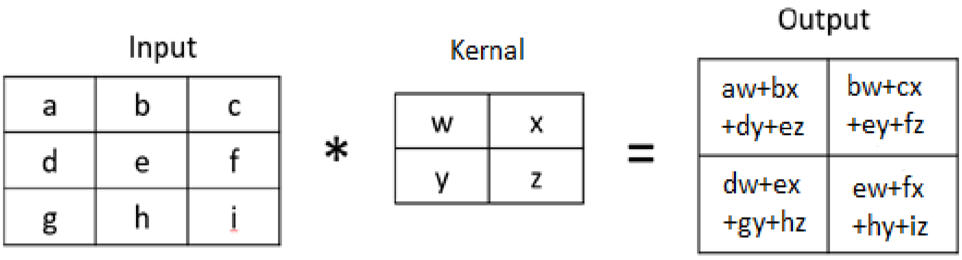

Figure 2.2. CNN convolution operation

Activation Layer

An activation function in a neural network defines how the weighted sum of the input is transformed into an output from a node or nodes in a layer of the network. In a deep learning neural networks, an activation function specifies how the counted weighted of the input is transferred into an output from a unit or units in a layer.

The activation layer function is often known as a "transfer function." The activation function is referred to as a "squashing function" if its output range is restricted. Most activation functions are nonlinear, so this "nonlinearity" in the layer or network configuration is described as "nonlinearity."

The activation function chosen does have a massive effect on the neural network's capabilities and efficiency, and different activation functions may well be utilized in different regions of the deep learning architecture.

Although networks are structured to use the similar activation function over all nodes in the models layer, the activation layer operation is applied inside or even after the internal operations of every node in the deep learning model (architecture).

Sigmoid

$\sigma(\mathcal{X}) = \frac{1}{1+e^{-x}}$

Leaky ReLU

$\max(0.1x, x)$

tanh

$\tanh(x)$

Maxout

$\max(\mathcal{W}_1^T \mathcal{X} + b1, \quad \mathcal{W}_2^T \mathcal{X} + b2)$

ReLU

$\max\{0, \mathcal{X}\}$

ELU

$$\begin{cases} \mathcal{X} & \mathcal{X} \geq 0 \\ \alpha(e^x - 1) & \mathcal{X} < 0 \end{cases}$$

Pooling layer

A popular strategy for arranging layers throughout a convolutional neural network is to add a pooling layer after the convolutional layer. Particularly, after adding a nonlinearity (e.g., ReLU) to the feature maps formed by a convolutional layer. This pattern can be reused one or more times in a particular model. This layer performs on every feature map independently to build a new collection of pooled feature maps with almost the same range of features.

Pooling is similar to applying a filter to feature maps in that it entails selecting a pooling procedure. The pooling layer operation or filter is probably smaller than the feature map; for example, it is always 2×2 pixels applied with 2 pixel strides.

This shows that the pooling layer is used to reduce the size of each feature map by a factor of two, i.e., each dimension will be halved, resulting in a feature map with one-quarter the number of values or pixels. A summary form of the features collected in the input is the outcome of applying a pooling layer and constructing down sampled or aggregated feature maps. Minor changes in the position of the feature in the input recognized by the convolutional layer result in a pooled feature map mostly with the feature in the same position, making this layer helpful. The model's variation to local translation is a feature added via pooling.

Fully connected layer

In a Fully Connected layer, neurons are completely connected to every activation with in prior layer, as the name implies. Layers that are fully connected have always been located at the network's end point. This layer receives an input volume (whatever the result of the conv, ReLU, or pool layer before it is) and outputs an N-dimensional vector, where N is the number of classes the program must realize. For example, in our experiment the number of classes N is equal to 2, the output will be Autistic (it means that the person or child is not normal), and the second output will be Non-Autistic (that means the person or child is normal).

Dropout

A training dataset with a little number samples is likely to quickly over-fit deep learning neural networks. Overfitting is known to be reduced by ensembles of neural networks with alternative model combinations, however training and maintaining many models adds to the computational cost and time. By randomly dropping out nodes during training, a single model can be utilized to simulate having a huge variety of distinct network models. Dropout is a regularization strategy that reduces overfitting and improves generalization error in deep neural networks of all types. It is computationally cheap and amazingly effective [27].

Huge neural networks trained on limited datasets have a tendency to over-fit the training data. When the architecture is tested on new data, such as a test dataset, the model learns the random variation in the training data, resulting in poor performance. Overfitting increases prediction error. Fitting all feasible distinct neural networks on the same dataset and averaging the predictions from each architecture is one way to reduce overfitting. In actuality, this is not possible, but it can be approximated using a small group of distinct models known as an ensemble. In a neural network, dropout can be applied per layer. Several types of layers, including dense fully connected layers, convolutional layers, and recurrent layers like the long short-term memory network layer, can be utilized with it. Dropout can be used on all or some of the network's hidden layers, and also the visible or input layer. but dropout cannot be used On the output layer, the dropout technique is shown below in Figure 2.3.

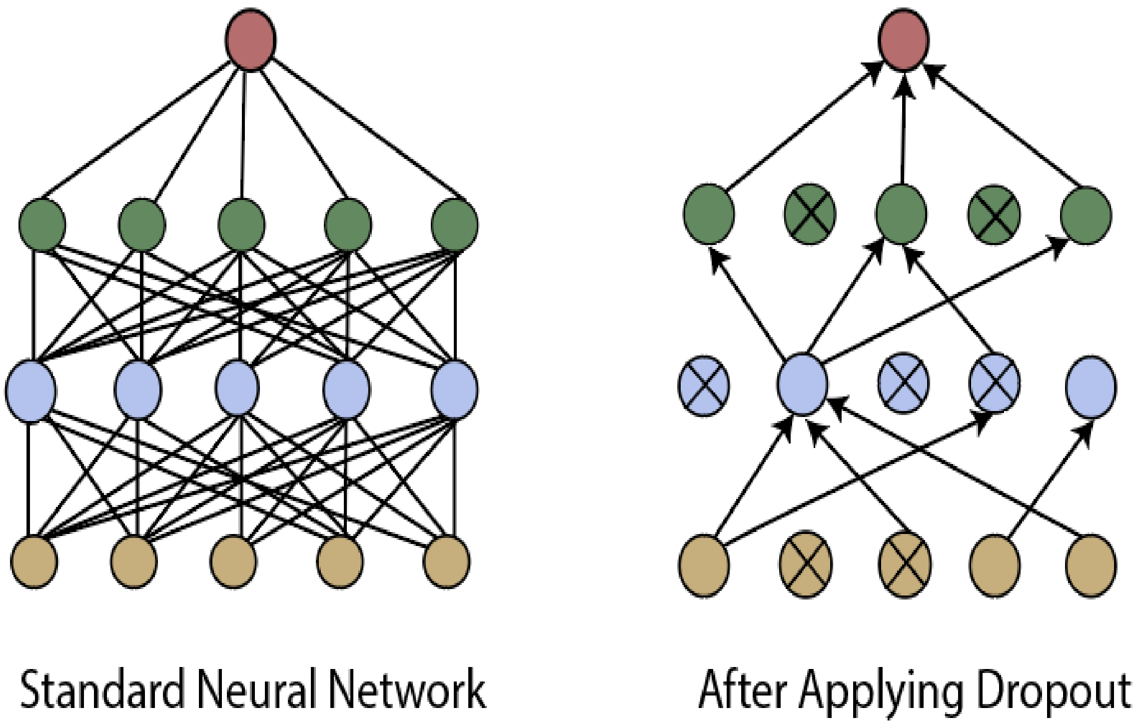

Figure 2.3. Dropout technique in Deep Learning

Batch Normalization

Since Deep learning neural network models with tens of layers are difficult to train because they are sensitive to the learning algorithm's randomly initialized weights and setup.

The balance of inputs to layers down in the network may shift after every mini-batch whenever weights are adjusted, which could be one cause for the difficulties. It could cause the learning process to follow a moving target indefinitely. The precise term for this shift in the distributions of inputs to layers in the architecture is "internal covariate shift" [28].

When we feed data into a machine learning or deep learning system, we usually modify the numbers to a balanced scale. Normalization is done in part to ensure that the models can generalize correctly.

Batch normalization is the process of putting additional layers to a deep learning neural network architecture to increase performance and make it more stable. The standardizing and normalizing procedures are performed by the new layer just on inputs of a prior layer [22].

The term "Batch" is used in batch normalization for a purpose. A typical neural network models are trained utilizing batch data, which is a collection of data input. Similarly, with batch normalization, the normalizing procedure is done in batches rather than as one input. The specific algorithm of Batch Normalization is shown below.

Input: Values of x over a mini-batch: $B = \{x_1 \ldots\ldots m\}$;

Parameters to be learned: γ, β

Output: $\{\Upsilon_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$

$\mu\mathcal{B} \longleftarrow \frac{1}{m}\sum_{i=1}^{m} x_i$ // mini-batch mean

$\sigma_{\mathcal{B}}^2 \longleftarrow \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu\mathcal{B})^2$ // mini-batch variance

$\hat{x}_i \leftarrow \frac{x_i - \mu\mathcal{B}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}}$ // normalize

$\hat{y}_i \leftarrow \gamma\hat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(\chi_i)$ // scale and shift

CNN architectures

Inception V3

Inception v3[29] is a convolutional neural network (CNN) model that was developed as a GoogleNet module to aid in image processing and object detection. That's the third version of Google's Inception Convolutional Neural Network, which was first shown off at the ImageNet Recognition Challenge. The goal of Inceptionv3's architecture was to enable deeper networks whilst keeping that model parameters low: it has "under 25 million parameters." Inception, like ImageNet, is a library of identified visual objects that aids object categorization in the field of computer vision. Many other applications have reused the Inceptionv3 model, which is frequently "pre-trained" via ImageNet. However, one application is in the field of life sciences, where that aids in the study of leukemia. Inception has a reduced computational costs than VGGNet or its higher-performing descendants. As a result, Inception architectures could then be used in big-data situations in which a massive volume of data performs tasks at low costs, or in circumstances where memory or processing power is intrinsically constrained, such as in mobile vision. Parts of these problems can be mitigated by using specialized solutions to targeted memory usage or by using computational methods to optimize the execution of specific tasks. These methods, on the other hand, add to the complexity. These techniques could also be used to improve the Inception design, extending the efficiency gap once again. Despite this, the Inception model's intricacy makes network upgrades more challenging. Large portions of the computing gains can be lost quickly if the model is ramped up rashly. Also, it doesn't give a good explanation of the elements that influence the GoogLeNet architecture's numerous design decisions. This makes it far more difficult to adapt it to new use-cases whilst keeping it efficient. For instance, if it is determined that significantly improving the performance of Inception-style architecture is essential, then expanding the number over all filter bank sizes will result in a 4x increase across both computation costs and parameters count. In many actual settings, this may be prohibitive or impractical, especially if the related rewards are minimal. By discussing a few broad concepts and optimizing ideas which have shown to be effective in scaling up convolutional networks. Our concepts are not confined to Inception-style networks, but they are simpler to detect in that context since the generic form of the Inception-style build pieces is flexible enough just to naturally accommodate those limitations. This is made possible by the Inception modules' extensive use of data reduction and paralleled architectures that enables for the mitigation of the effect of structural changes on surrounding components. However, one must exercise caution in doing so, as certain fundamental precepts must be followed in order to preserve good model quality.

DenseNet121

DenseNet121[30] is a convenient and straightforward model that is built by alternately adding dense block and pooling layers. A straight feed-forward link is made of each layer to every other layer in the DenseNet121 design. The DenseNet121 contains N(N1)/2 direct connectivity between every consecutive layer, while a conventional convolutional network having N layers has N connections between each following layer. The DenseNet121 uses a lower connection between the first last layers, which improves the architecture's accuracy and efficiency significantly. The feature maps from every layer are concatenated with the preceding one as the input progresses through the successive layers. By connecting directly every layer for all layers preceding it, the DenseNet model eliminates the requirement for feature replication and encourages feature reuse. While dense connection is more economical, we believe it introduces redundancy when early characteristics aren't needed in later levels. This allows following layers to access feature maps from any layer. When compared to a standard convolutional neural network, the number of parameters is less. The DenseNet121 solves the issue of vanishing gradients and provides feature reuse. However, for a smaller dataset, the model's regularization effect decreases overfitting to a higher extent, as illustrated.

ResNet50

A well-known artificial neural network (ANN) is the residual neural network, also known as ResNet [31]. It arranges using pyramid cell constructions from the cerebral cortex. The residual neural networks achieve this by moving between layers utilizing shortcuts or "skip connections." Traditional residual neural network architectures with two or three layers of skipping, batch normalization, and nonlinearities in between are used by experts. In some circumstances, data scientists also use an additional weight matrix to learn the skip weights. "Highwaynets" is the term that describes this situation. "Densenets" are models with many parallel skips. When discussing residual neural nets, non-residual nets are often known as plain networks. Avoiding fading gradients and other difficulties is a major reason for skipping layers. Because the gradients are back-propagated to prior layers, the gradient may become incredibly tiny as a result of this recurrent process. The majority of people do this by using the activations from previous layers till the next one learns in certain weights. These weights adjust to the upstream layers during training and emphasize the layer that was skipped previously. The weights used to link the adjoining layers play a role in the most basic situation.

This, however, just works when most of the intermediate layers are either linear or overlay the non-linear layer. If this is not the case, using a separate weight matrix for missed connections would be beneficial. In such instances, it would be preferable if you explored using Highwaynet. Within the first training step, skipping layers removes complexity from the network, making it simpler. It tenfold learning speed while reducing the influence of vanishing gradients. Why? There aren't many layers to spread through since there aren't many to begin with. The network then gradually restores the expert layers when learning the feature space. They grow closer to the manifolds and understand things more quickly as the training progresses, and each layer expands. A neural network with no residual parts has much more ability to experiment the feature space, which makes it vulnerable to perturbations, forcing it to depart the manifolds, and necessitating the recovery of the additional training data.

Xception

Xception [32] is a Depth-wise Separable Convolutions-based deep convolutional neural network model. Google expert’s researchers came up with the idea. Inception modules in convolutional neural networks are described by Google as an interim step between basic convolution and the depth-wise separable convolution process (a depth-wise convolution followed by a pointwise convolution). Throughout this sense, a depth-wise separable convolution can be thought of as an Inception module with the most towers possible. As a result of this discovery, they propose a new deep convolutional neural network design based on Inception. where depth-wise separable convolutions have swapped Inception modules. Xception is a well-designed model that is based on two fundamental principles: first, Depth-wise Separate Convolution, and the second is abbreviations between Convolution blocks as in ResNet. Depth-wise Separable Convolutions are a type of convolution which is claimed to be substantially faster to compute than traditional convolutions. Depth-wise separable convolutions have just been invented to reduce the cost of these kind of processes. They are grouped into two main sections: Depth-wise Convolution and Pointwise Convolution. In the implementation, Xception uses a structure made up of Depth-wise Separable Convolution blocks with Maxpooling layers, all of which are linked by shortcuts similar to ResNet implementations. Xception specificity is the Depth-wise Convolution is not preceded by either a Pointwise Convolution in Xception; rather, the sequence is reversed.

MobileNet

Sandler, Howard (2018) invented MobileNet model [33], that achieves a decent compromise between efficiency and computationally cost. MobileNet used separable convolution and Depth-wise convolution to significantly lower convolutional layer parameters while keeping high classification efficiency. Residual bottleneck blocks were recommended in MobileNet for reusing the feature maps. As a result, MobileNet can provide high classification accuracy with significantly shorter training time. The MobileNet architecture is constructed on depth-wise separable convolutions, that were a type of factorized convolution that divides a standard convolution into a depth-wise convolution and a 1x1 pointwise convolution. The depth-wise convolution used by MobileNet offers a distinct filter per each input channel. The depth-wise convolution's outputs are then combined using an 1x1 convolution even through the pointwise convolution. Inside one stage, a conventional convolution filters and merges inputs to develop a new set of results. The depth-wise separable convolution divides it into two layers: one for filtration and the other for combination. This factorization reduces computing time and model size significantly.

Besides its initial layer, which is a full convolution, the MobileNet structure is building on depth-wise separable convolutions. We may quickly explore network structures to identify a suitable network by specifying the network in these basic terms. Except for the final fully connected layer, that has no non-linearity and feeds this into softmax layer for classifying, all layers of the model are followed by batch-normalization and ReLU non-linearity. With depth-wise convolution, 1x1 pointwise convolution, and batch-normalization and ReLU after every convolutional layer, batch-normalization and ReLU non-linearity is added to the factorized layer. In the depth-wise convolutions and also the first layer, down sample is performed via stride convolution. Even before fully connected layer, a last pooling layer decreases the spatial resolution to 1. Depth-wise and pointwise convolutions are counted as independent layers.

Support Vector Machines

Vladimir Vabnik and Alexey Shervonenikis devised the support vector machine (SVM) algorithm through 1963 [34]. The accepted algorithm has been developed by Corrina Cortes and Vabnik in 1993 and released in 1995. The support vector machine (SVM) algorithm, also known as the support vector network (SVN), is a data analysis algorithm that uses regression and classification processes to analyze data. The SVM algorithm's popularity in the machine learning and data mining sectors stems from its simplicity in terms of comprehension and application. The term "support vector machine" is a great concept for just a machine that analyzes and predicts data. In truth, this moniker is a good fit for its formidable capabilities, which have had a lot of success, particularly in the information classification phase. The SVM algorithm's operating principle is to delimit a border for a related point field (belonging to a particular class). It is required to examine whether the samples (test samples) falls inside the border when a limitation is drawn (on the training samples) for any additional points (test sample) to be categorized. The training data set (with the exception of data extremely close to or situated on the border) are regarded redundant just at the time of the border creation, so the technique does not depend on them until it completes its task. This makes it perform better than most other algorithms methods available in the data classification stage. It only requires a basic set of points to clearly identify and determine boundaries. These data points are referred to as "support" since they effectively support the boundary. The name "vector" comes from the fact that each item is a vector: each line of data contains info for a collection of attributes. The classifier displays as a single direction or a curve in basic scenarios when the data is two aspects, as well as when the data on both sides is in the case of multiple classes, each with its own set of attributes. When dealing with three-dimensional data, the classifier displays as a complicated surface or a plane when the data is classified linearly. More than three-dimensional data, on the other hand, is extremely difficult to perceive. As a result, the classifier has become a generic name for the boundaries that divide these various types of data. There are a number of feature sets visible. Is there a classifier that is better than the others? The right solution would be that the classifier split the classes entirely and appropriately. In other words, the optimum border line is a classifier with just an equal spacing (maximum distance) between every region of data from both sides. The margin is indeed the maximum distance between the border and the nearest patterns on the both sides of a classifier.

3. MATERIAL AND METHOD

Dataset

Dataset used for our experiment that were from a Kaggle datasets site [35] that has over three thousand images of children's faces, and this dataset images include of both autistic and non-autistic children. Because the author just had access to websites to obtain all of the images, this dataset is a little unusual. The data can be obtained in two formats: separated into training, testing, and validation or combined. The offered breakdown of training, testing, and validating subgroups will be handy if we decide to develop our own deep learning model. The validation component will be critical in evaluating the model's quality, which indicates we won't have to rely solely on the architecture's accuracy to assess its quality. The sections of training, testing, and validating are even further divided into autistic and non-autistic folders. The autistic training folder has 1268 facial photos, while the non-autistic training directory has the same number of images. There are 150 photographs in each of the autistic and non-autistic testing folders, for a total of 300 images. Finally, there are 100 images in the validation category: 50 non-autistic and 50 autistic facial images. that could use an existing model, it could use the aggregated pictures as it will be able to handle the quantity of data required for training and testing. As seen below in Figure 3.1, Autistic and Non-Autistic samples of the dataset.

Figure 3.1. Autistic and Non-Autistic samples

Data Augmentation

Data Augmentation [36] is a strategy or technique to artificially raising the amount of a dataset by appending certain modifications to an original dataset and then combining it with the original dataset to produce 'slightly modified and multiplied' data. Data augmentation is the process of taking all or some of the items in a dataset and modifying them numerous times in different ways to create a larger dataset.

In deep learning models, collecting and categorizing data is a time-consuming and costly operation. Data augmentation can produce in datasets that help clients save money on operations. It also handles the problem of restricted dataset size and data fluctuation at the same time. This enhances the model's overall efficiency in various of cases. The first step in the deep learning process is to produce or obtain the necessary dataset for training the deep learning model. Deep learning models are capable of recognizing the items for which they have been programmed. However, if they are not part of the training, they will struggle to deal with the many scenarios. If the network is trained with all of the training photos oriented in one way, it may be unable to detect the objects in horizontally and/or vertically flipped images. The cause for this is that, despite belonging to the same object, the characteristics it outputs differ from the ones it learned during training. The availability of high-quality data is often a concern in most cases. It might only be available in modest amounts or just not. It would be difficult to collect a dataset large enough to attain the requisite accuracy in such circumstances. Underfitting or overfitting can occur if the dataset is not accessible in sufficient quantity or with the necessary level of variance [37].

The deep learning architectures in the discriminative tasks have made amazing progress. Deep learning models, sophisticated processing, and access to massive data have all contributed to this. Due to the improvement of convolutional neural networks, deep neural networks are widely implemented to Computer Vision applications such as image classification, object detection, and image segmentation (CNNs). The spatial properties of images are preserved using parameterized, sparsely connected kernels in these neural nets. Convolutional layers down sample images' spatial resolution while growing the depths of their feature maps in a sequential manner. This set of convolutional transforms can produce considerably lower-dimensional but more usable image meanings than could be achieved by hand. The success of convolutional neural network models has piqued interest in using Deep Learning to solve computer vision problems. There are many types of data augmentation but images type is most important and well-known of data augmentation and entails transforming photos in the training dataset into transformed copies that are of the same classes as the main image [36], [37].

The transformations processes refer to a set of techniques used in the field of image processing., such as shifts, cropping, zooming, rescale, flips, random rotating, and much more. The goal is to add additional, believable samples to the training collection. This refers to changes of the training classes images that the architecture is likely to observe. A horizontal flip of a cat shot, for example, may make sense because the snap might have been taken from either the left or right. A vertical flip of a cat photo makes little sense and is unlikely to be acceptable, considering that the model is unlikely to view an upside down cat image.

As a result, it is evident that the precise data augmentation methods utilized for a training dataset should be selected carefully, taking into account the training dataset as well as understanding of the issue area. In furthermore, a small prototype dataset, models, and training run can be used to evaluate data augmentation methods alone and together to evaluate if they result in a significant improvement in deep learning architecture performance.

Advanced deep learning methods [37], including the convolutional neural network (CNN), can discover features that are independent of where they appear in the image. However, augmentation can help with this transform invariant method to learning by assisting the models in learning features which are also transformation consistent, including left-to-right to top-to-bottom ordering, light levels in images, and so on. Generally, image data augmentation is used only just on training dataset, not the test or validation datasets. Data preprocessing, such as image resizing and pixel scaling, differs in that it must be done effectively throughout all datasets that interacts with the architecture. In the Figure 3.2. the rotating of the images technique shown below.

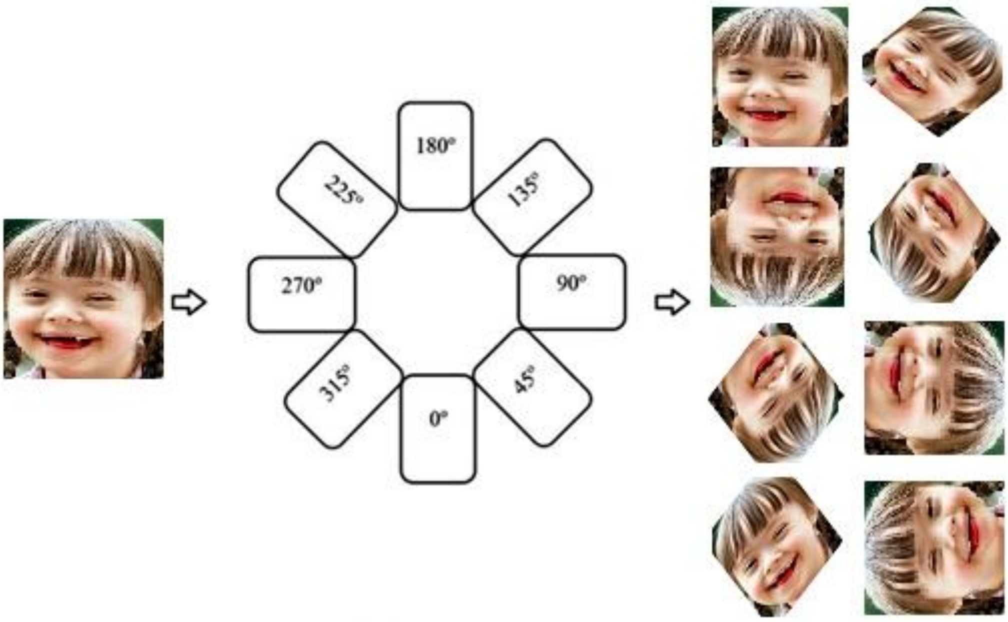

Figure 3.2. Procedure of rotating for image at various angles (Rotation Transformation)

Transfer learning

Transfer learning is that once a machine learning model that has already been trained is used to solve a separate but related issue. For example, if we trained a basic classifier to predict if an image contains a backpack, we might utilize the model's information to distinguish other things like sunglasses.

The main concept is to apply what a model has learned from an issue with a large amount of labeled training data to a new issue with less data. Rather than starting from zero, we begin using patterns gained by completing a related task. Because of the large volume of CPU power required, transfer learning is typically employed in computer vision and natural language processing tasks like sentiment analysis [38].

Transfer learning is more of a "design methodology" within the subject of active learning than it is a machine learning technology. It's also not a stand-alone component or topic of research in machine learning. Nonetheless, it has gained a lot of traction when used in conjunction with neural networks, which demand a lot of data and processing power. Transfer learning is when a machine learning model that has already been trained is used to solve a separate but related issue. For example, if you trained a simple algorithm to predict if a data contains a backpacks, then might utilize the model's information to distinguish other things like sunglasses.

The aim to use what that have learned in one problem to boost generalizations in another through transfer learning. the weights learned by a deep learning architecture at "task A" were moved to a new "task B".

In computer vision, neural nets typically focus on detecting edges throughout the first layer, forms in the middle layer, and task-specific properties in the latter layers. The early and intermediate layers are utilized in transfer learning, and the subsequent layers are only retrained. It makes use of the labeled data from the task it was trained on.

The strive to transfer much more required knowledge from the prior issue that the model was trained over to the current issue at hand in transfer learning. Depending on the situation and the data, this knowledge might take many different forms. It could, for example, be the way models are built, which makes it easier to recognize new items [39].

The transfer learning [11] provides a number of advantages, the most important of which are reduced training time, improved neural network performance (in most circumstances), and the absence of a large amount of data. To train a deep learning model, a large volume of data is generally required, but accessing to that data isn't always possible – this is when transfer learning gets in aid. Because the network has already been pre-trained, a good machine learning architecture can be generated with relatively little training data using transfer learning. This is especially useful in natural language processing, where huge labelled datasets require a lot of expert knowledge. Additionally, training time is minimized because building a deep learning network from start on a challenging job can takes days or even weeks.

Two methods of implementing transfer learning have been used to handle computer vision issues.

- First, if you have a large collection of images (>1,000 per class), you could use weights from a model trained on a separate dataset to begin a new architecture. During training, you'll fix a few layers (usually the initial convolutional layers) while optimizing the parameters of the higher-level layers. The idea is to limit the number of factors that must be tuned while keeping the lower-level layers in use. Lower-level layers are likely to be comparable regardless of the problem domain, and the model has the flexibility to combine higher-level levels as needed to solve the problem.
- Second, if just a limited number of images are provided (less than 1,000), retraining an existing network will almost always result in overfitting. When compared to the quantity of photos, the number of variables that need to be optimized is simply too huge. However, a large pre-trained model (trained on that huge dataset) could be employed as a feature if the input was similar in appearance to the images in that huge dataset. You delete the last N levels of a large network (usually N=1 or N=2), in particular. And using the output of the large pre-trained model as a feature representation of the images data. This is predicated on the assumption that perhaps the pre-trained model's first layers learn issue features. These features could then be combined with classic computer vision techniques such as SVMs or logistic regression. The pre-trained model is used as a feature rather than having to identify the features. Transfer learning process explained below in Figure 3.3.

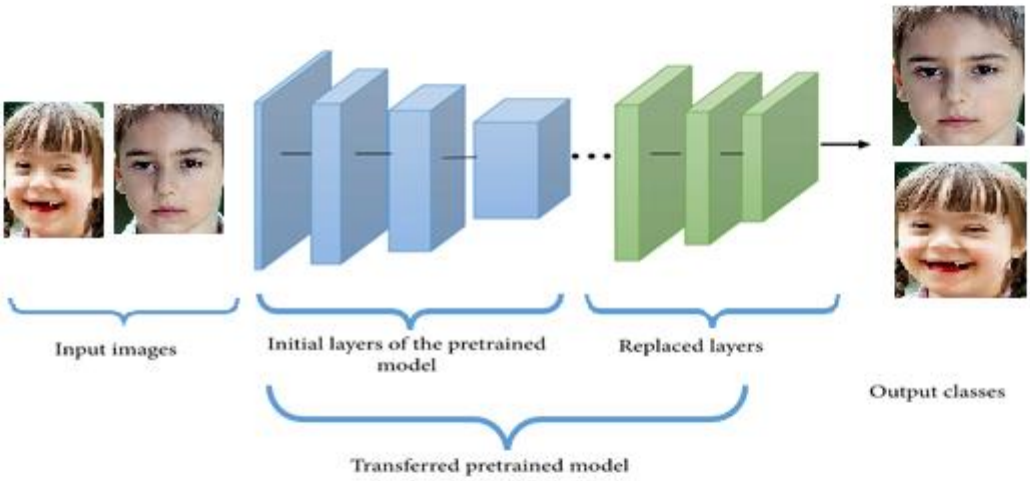

Figure 3.3. Pre-trained models are used in the transfer learning procedure.

Evaluation metrics

Image classification and object detection are examples of deep learning techniques that are often employed. It's critical to choose the correct evaluation metrics when deciding which model to be using, how to adjust the hyper parameters, whether regularization methods are needed. The most crucial step in the development of whatever deep network is to assess its efficiency. As a result, the topic of how to evaluate the success of a deep learning model emerges [40]. Then how could we determine when to calling it what it is, when and where start and end the training and evaluation?

Deep learning tasks are exposed to evaluation measures. Different metrics exist for classification and regression analysis. Some measures, such as precision-recall, are valuable for a variety of purposes. Supervised learning, which accounts for the vast majority of deep learning apps, includes classification and regression. It must be able to increase our model's overall predictive capacity using several metrics for measuring performance before deploying it on unknown data. When the appropriate architecture is deployed on unseen data, failing to properly evaluate the deep learning architecture using various evaluation measures and relying solely on accuracy might lead to problems and incorrect predictions. Classification metrics for image classification. There are two types of image classification tasks: binary and multi-classification. There are just two prediction classes in binary classification. Examples of binary classification include the detection of diseases such as autism, Malaria, or cat/dog classifying, among others. There are more than two classes of prediction in multiclass classification; examples of Multi-label classification are MNIST, CIFAR, and others.

The accuracy [29] Accuracy is usually the first parameter used in classifications. It's a straightforward statistic that calculates the proportion of correct to incorrect predictions. however, always valid. or it is a measure of a network's capacity to predict outputs. Where the overall number of right predictions is divided by the entire number of predictions. The prediction is rounded to zero or one according on its value, and then compared to the genuine label. It is a correct prediction if the predictions and the label are equal. The average accuracy is usually shown as a function of epochs, and the plot gives a summary of the average accuracy throughout the training, below the accuracy mathematical equation.

$$\text{Accuracy} = \text{True Positive} + \text{True Negative} / \text{True Positive} + \text{False Positive} + \text{False Negative} + \text{True Negative}$$

The sensitivity (recall) is The number of cases that get a positive outcome on just this test among those who genuinely have the ailment (as evaluated by the 'gold standard') is referred to as sensitivity (true positive rate).

Sensitivity is a measurement of how successfully a test accurately identify true positives in a detecting test. There is always a trade-off between specificity and sensitivity in all diagnosis and treatment testing, with higher sensitivities implying lower specificities and inversely. The frequency of false negatives must be limited if the purpose of the test is to identifying those who have an illness, which necessitates high sensitivity. That is, those with the ailment should have a high chance of being detected by the test. This is especially crucial when the consequences of not treating the ailment are severe and/or the medication is highly efficient with little adverse effects. Below the sensitivity mathematical equation.

$$\text{Sensitivity} = \text{True Positive}/\text{True Positive} + \text{False Negative}$$

Precision: The proportion of true positives (i.e., total number of items that correctly categorized as belongs to the positive category) calculated by dividing the number of objects classified as belongs to a positive class is the category accuracy in the classification task (i.e., the total number of all true positives and false positives, as these are elements which have already been categorized mistakenly as belonging to the class). Precision defined in the below mathematical equation.

$$\text{Precision} = \text{True Positive}/(\text{True Positive} + \text{False Positive})$$

The fraction of actual negatives that were anticipated as negatives is known as specificity (or true negative). This means that a part of true negatives will be forecasted as positives, which may be referred to as false positives. This percentage is also known as the false positive rate. The summation of specificity and false positive rate is always 1 in this case. The percentage of individuals not afflicted from the diseases who were accurately predicted as not afflicted with the disease is known as specificity. To put it another way, the individual who is healthy was anticipated to be healthy because of specificity. The specificity is defined in the below mathematical equation.

$$\text{Specificity} = \text{True Negative} \,/\, \text{True Negative} + \text{False Positive}$$

The percentage of the right classifications of the autistic or ill person, where TP stands for genuine positive. Of the right classifications of the autistic or unhealthy person, where TP stands for true positive. The TN stands for true negative, and it refers to the number of times a non-Autistic or normal person has been correctly classified. The FP stands for false positive, which refers to the number of autistic people who have been categorized as non-autistic. The FN stands for false negative, which refers to the number of non-Autistic people who have been labeled as Autistic [5]. TP, TN, FP, and FN are commonly used to create a confusion matrix, which is specified in the mathematical equation below. A better classifier is indicated by greater detection accuracy values.

$$Confusion\,matrix = \begin{pmatrix} \mathrm{TN} & \mathrm{FP} \\ \mathrm{FN} & \mathrm{TP} \end{pmatrix}$$

The F-score or F-measure is a measurement for determining how accurate a test is. It is calculated using the test's precision and recall, with precision equaling the number of true positive results divided by the total number of positive outcomes, which include those that were mistakenly classified, and recall equaling the number of true positive outcomes divided by the total number of items that should have been defined as positive. In diagnostics binary classification, precision is also referred to as positive predictive value, while recall is also referred as sensitivity. It is defined in the mathematical equation below.

$$\mathrm{F1-score} \;=\; 2 * (\mathrm{Precision} \;*\; \mathrm{Recall}) \,/\, (\mathrm{Precision} \;+\; \mathrm{Recall})$$

Hardware and software configuration

Hardware configuration

All experiments in this thesis were carried out on a PC with an i5 processor., 4 GB memory, and Windows 10 64-bits. In addition, The Scientific Python Development Environment, or Spyder, is just a free integrated development environment (IDE) that comes with Anaconda. It operates properly without the requirement for administrator credentials and can quickly deal with huge data computation. The network architecture's performance was evaluated using Spyder software. It provides a framework for designing and implementing CNNs, including apps and visualizations to view network activations and track network training progress. Table 4.1 contains all of the pertinent information.

Table 4.1. Hardware identifications

Hardware	Characteristics
Memory	4 GB
Processor	Intel Core i5-3230M CPU @ 2.60 GHz
Graphics	Intel® HD Graphics 4000
Operating system	Windows 10, 64 bits

Software configuration

In Several improved approaches are frequently employed to allow faster the evaluation of a deep neural network, which is based on the classic small batch gradient descent methodology. The following two characteristics of standard enhancement techniques are primarily improved:

Learning rate reduction and gradients direction optimization, while learning rate attenuation is defined as modifying the parameter updating step size, and gradient direction optimization is defined as modifying the updated direction. Batch or stochastic gradient descent techniques could also benefit from these enhanced optimization strategies. Furthermore, the learning rate used to be 0.0001, the highest numbers of epochs were 35, the Validation Frequency is 3, and we shuffled the dataset each epoch. Last, the CNN architectures have all been trained using a 32 batch size. The specifics are shown in Table 4.2.

Table 4.2. CNN network's training option hyper-parameters and values

Training option Parameters	**Value**
Optimization algorithm	Adam
Learning rate	0.0001
Epochs	35
Shuffle	Every epoch
Batch size	32
Validation Frequency	3

4. RESULT AND DISCUSSION

Experimental results

In this thesis, the challenge of classifying autism via medical images was evaluated using state-of-the-art pre-trained networks. The goal of this study was to evaluate CNN models and assess their accuracy, precision, sensitivity, specificity and F1-Score by fine-tuning. And then in another time we augmented our dataset, and then we combined CNN pre-trained models with support vector machine (SVM) classifier and tried to classification process on the dataset. The results are shown in Table 4.3.

Table 4.3 Pre-trained networks' results

	With SVM					**Without SVM**				
Performance Measures	**MobileNet**	**Densenet121**	**Inception V3**	**Xception**	**Resnet50**	**MobilNet**	**Densenet121**	**Inception V3**	**Xception**	**Resnet50**
Accuracy	**98.83**	93.05	92.50	92.00	75.60	95.75	91.96	91.46	90.60	72.41
Recall	**97.33**	85.50	86.05	88.40	77.60	91.33	83.33	83.33	86.66	74.00
Precision	**98.64**	86.00	87.50	88.00	79.55	89.54	84.45	85.61	85.52	75.51
Specificity	**97.36**	85.75	86.75	88.75	78.00	91.15	83.55	83.76	86.48	74.50
F1-score	**97.98**	85.74	86.76	88.19	78.56	90.42	83.88	84.45	86.06	74.74
Time (min.)	**327**	370	335	599	421	90.5	330.8	239.7	445.4	346.1

The pre-trained networks all performed similarly and statistical significance. To begin, consider the accuracy matrix. The best results we achieved in this study were using MobileNet with Support Vector Machine classifier. By gaining 98.83 accuracy in results in (327 minutes). and then followed by the other architectures (MobileNet + SVM, Densenet121 + SVM, Inception V3 + SVM, Xception + SVM, and ResNet50 + SVM). with 95.75-percent, with 91.96 percent, 91.46 percent, 90.60 percent, and 72.41 percent. and also while performing the pre-trained models without using SVM classifier. MobileNet achieved best accuracy among other pre-trained models.

MobileNet with SVM, MobileNet, Densenet121, InceptionV3, Xception, and ResNet-50 The confusion matrix of the pre-trained architectures with both the best possible result based on the evaluation measurements MobileNet with SVM, MobileNet, Densenet121, InceptionV3, Xception, and ResNet-50 is also included. The performance of the classifier may be visually evaluated based on the results, and this can be determined which subclasses are highlighted by the neurons of the pre-trained architectures. These rows of the confusion matrix belong to the output classes, whereas the columns belong to the actual classes. As shown in the confusion matrix in Figure 4.1. the

MobileNet with Support Vector Machine (SVM) classifier network properly classified 146 of 150 standard images, Furthermore, it approximated four conventional photos as Autistic, while properly identifying 148 of 150 autistic images and incorrectly predicting two.

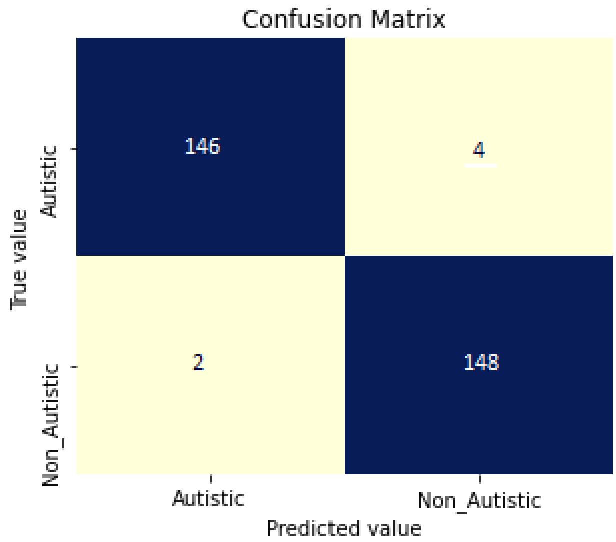

Figure 4.1. The Confusion Matrix of MobileNet with SVM classifier.

As shown in Figure 4.2, the Learning Curve of MobileNet with SVM classifier.

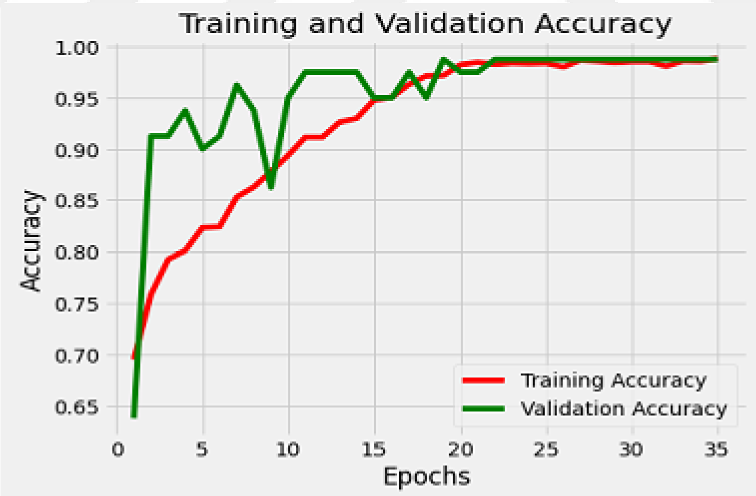

Figure 4.2. The Learning Curve of MobileNet with SVM classifier.

As shown below in Figure 4.3, the Loss Function of MobileNet with SVM classifier.

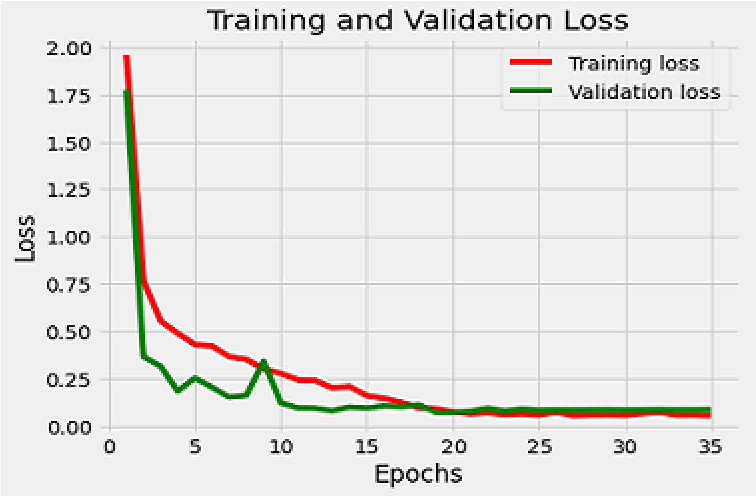

Figure 4.3. The Loss Function of MobileNet with SVM classifier.

As shown in the confusion matrix in Figure 4.4. the MobileNet model properly classified 137 of 150 standard images, Furthermore, it approximated 13 conventional images as Autistic, while properly identifying 134 of 150 autistic images and incorrectly predicting 16.

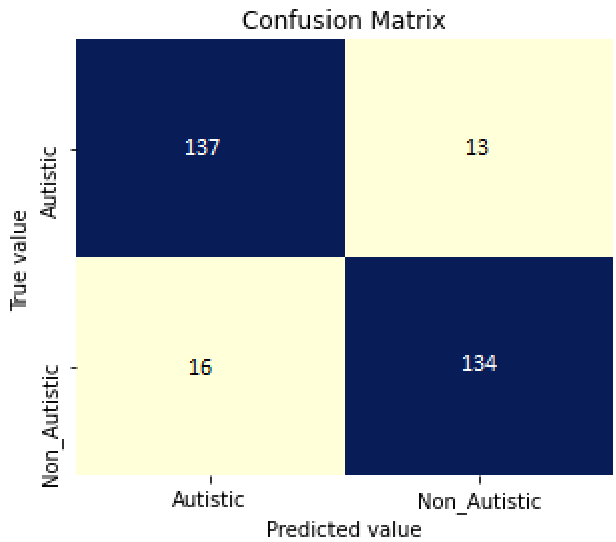

Figure 4.4. The Confusion Matrix of MobileNet model.

As shown below in Figure 4.5, the Learning Curve of MobileNet model.

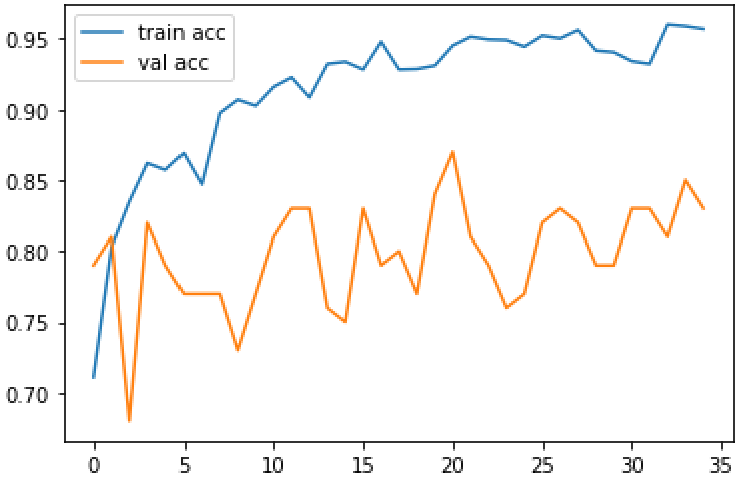

Figure 4.5. The Learning Curve of MobileNet model.

As shown below in Figure 4.6, the Loss Function of MobileNet model.

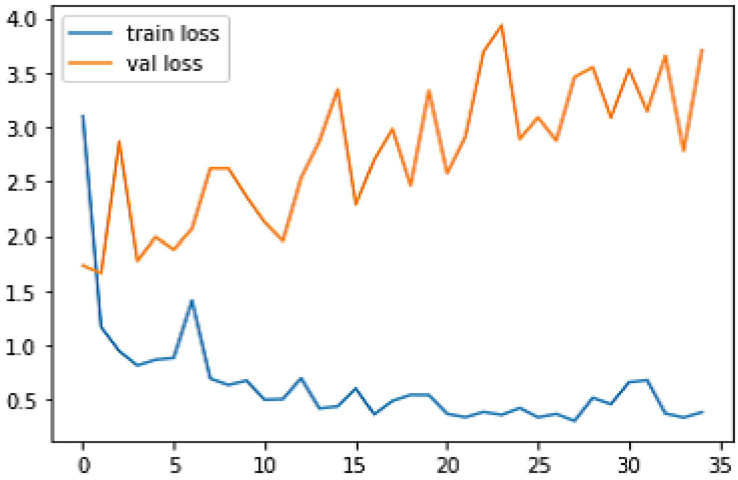

Figure 4.6. The Loss Function of MobileNet model.

As shown in the confusion matrix in Figure 4.7. the Densenet121 model properly classified 125 of 150 standard images, Furthermore, it approximated 25 conventional images as Autistic, while properly identifying 127 of 150 autistic images and incorrectly predicting 23.

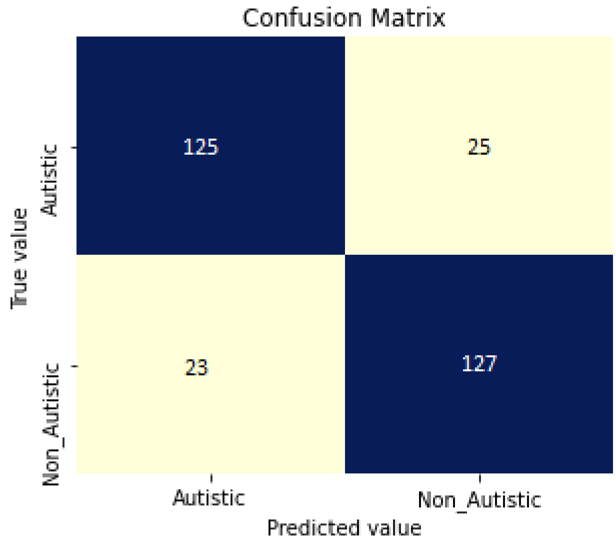

Figure 4.7. The Confusion Matrix of Densenet121 model.

As shown below in Figure 4.8, the Learning Curve of Densenet121 model.

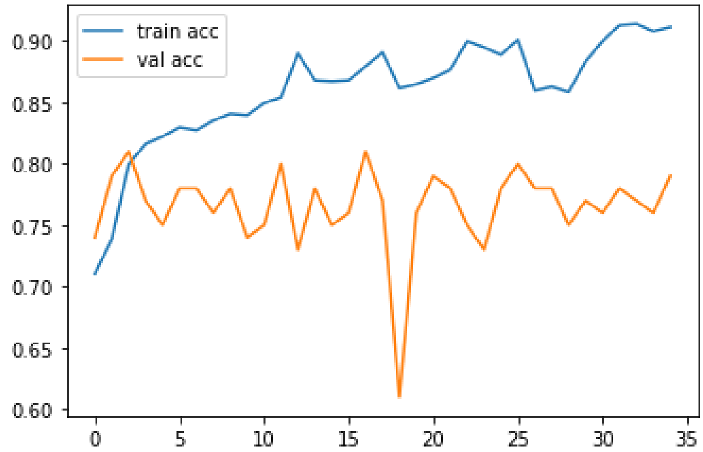

Figure 4.8. The Learning Curve of Densenet121 model.

As shown below in Figure 4.9, the Loss Function of Densenet121 model.

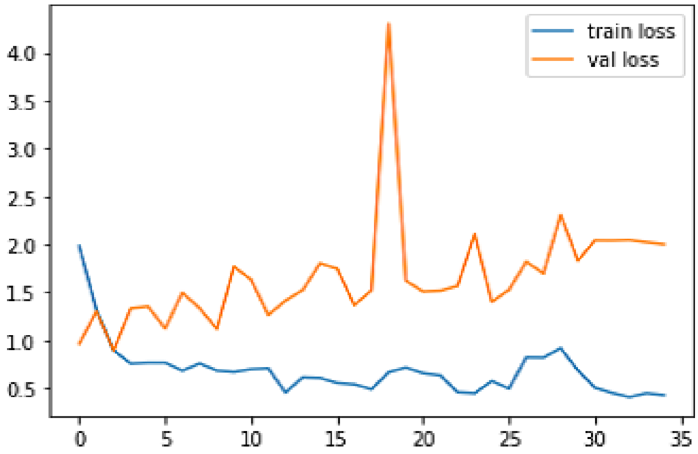

Figure 4.9. The Loss Function of Densenet121 model

As shown in the confusion matrix in Figure 4.10. the Inception V3 model properly classified 125 of 150 standard images, Furthermore, it approximated 25 conventional images as Autistic, while properly identifying 129 of 150 autistic images and incorrectly predicting 21.

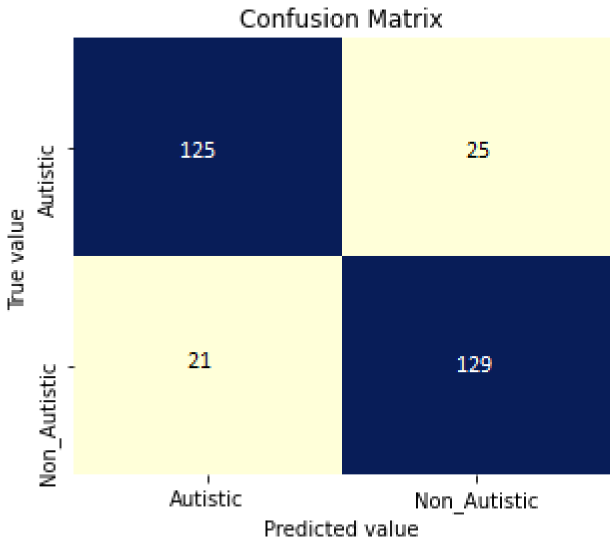

Figure 4.10. The Confusion Matrix of Inception V3 model.

As shown below in Figure 4.11, the Learning Curve of Inception V3 model.

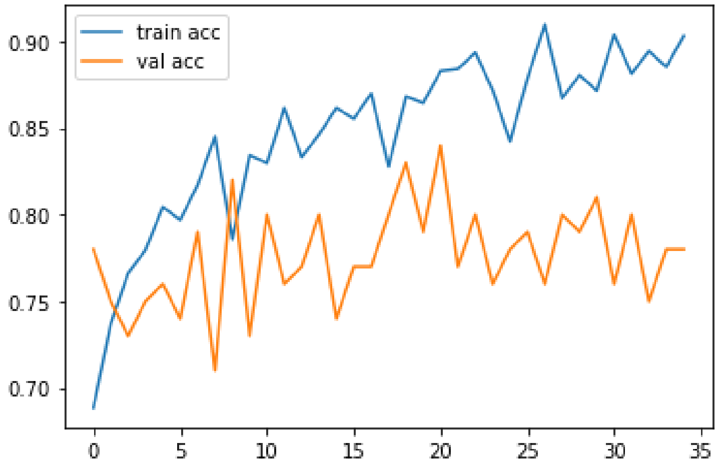

Figure 4.11. The Learning Curve of Inception V3 model.

As shown below in Figure 4.12, the Loss Function of Inception V3 model.

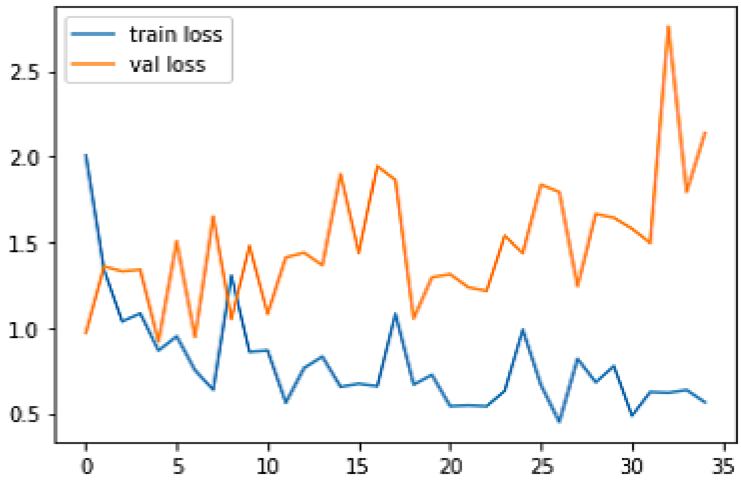

Figure 4.12. The Loss Function of Inception V3 model.

As shown in the confusion matrix in Figure 4.13. the Xception model properly classified 130 of 150 standard images, Furthermore, it approximated 20 conventional images as Autistic, while properly identifying 128 of 150 autistic images and incorrectly predicting 22.

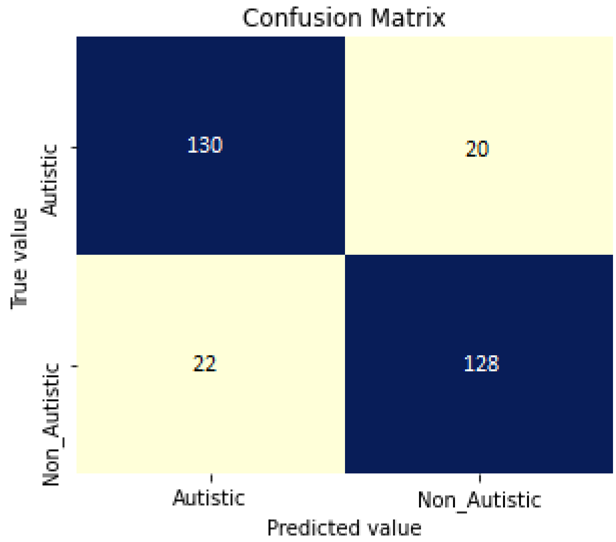

Figure 4.13. The Confusion Matrix of Xception model

As shown below in Figure 4.14, the Learning Curve of Xception model.

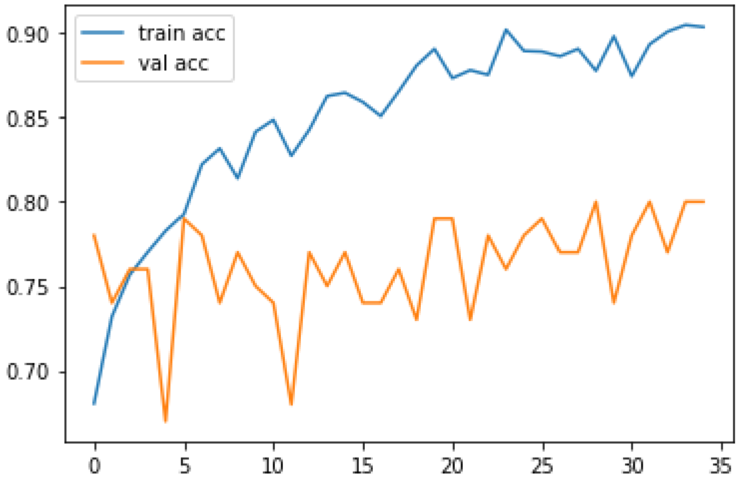

Figure 4.14. The Learning Curve of Xception model.

As shown below in Figure 4.15, the Loss Function of Xception model.

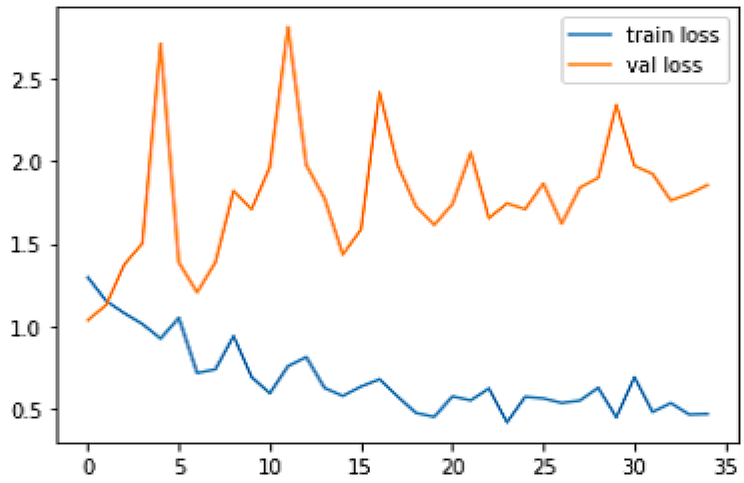

Figure 4.15. The Loss Function of Xception model.

As shown in the confusion matrix in Figure 4.16. the ResNet50 model properly classified 111 of 150 standard images, Furthermore, it approximated 39 conventional images as Autistic, while properly identifying 114 of 150 autistic images and incorrectly predicting 36.

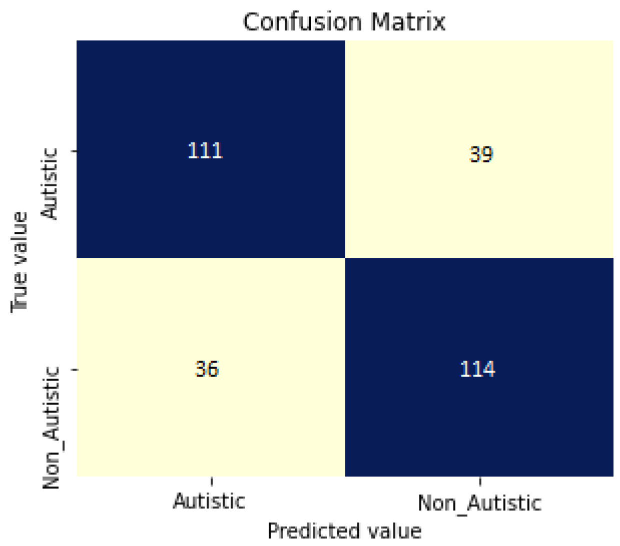

Figure 4.16. The Confusion Matrix of ResNet-50 model

As shown below in Figure 4.17, the Learning Curve of ResNet-50 model.

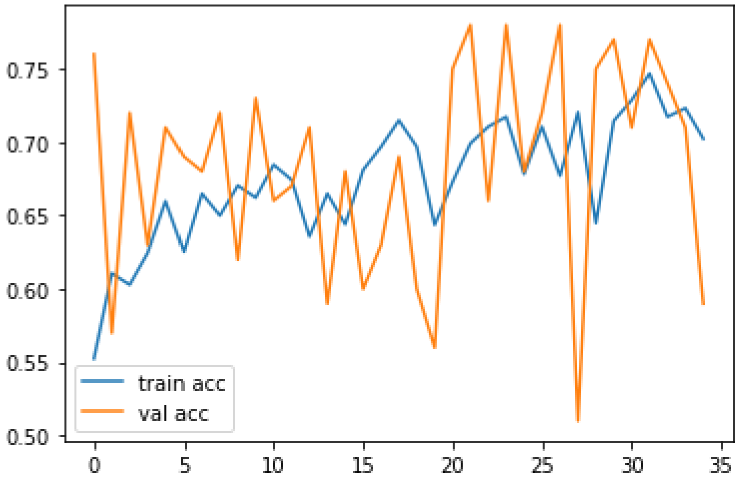

Figure 4.17. The Learning Curve of ResNet-50 model.

As shown below in Figure 4.18, the Loss Function of ResNet-50 model.

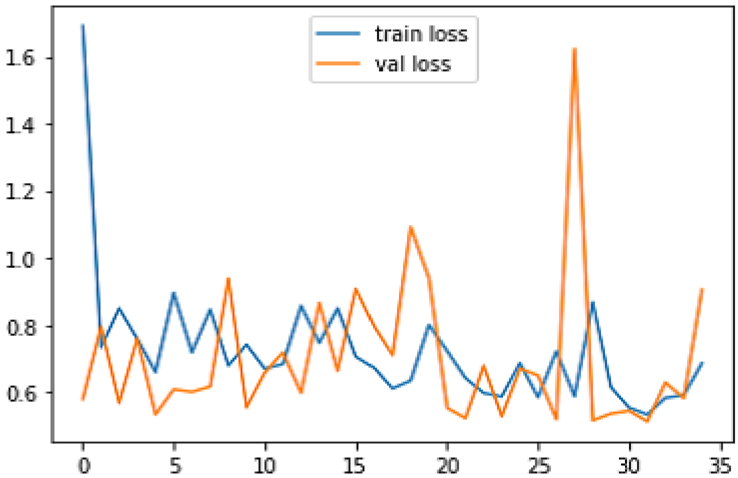

Figure 4.18. The Loss Function of ResNet-50 model.

Discussion

Early detection of autism is very important for the development of the affected child, so the development of these classifications based on a set of data related to it is a reason for early detection of autism, and this enhances the adaptation of the patient to normal life. The dataset can be used in two different ways, both of which are popular for deep learning tasks. The dataset is divided into Training, Test, and Validation sets, which is a normal procedure. The train is the title of the training set. That consists of two subdirectories known as (Autistic & Non-Autistic) [12], [13]. There are 1468 facial pictures of children with autism in the Autistic subdirectory in 224 X 224 X 3, jpg format. And there are also 1468 photos of children who do not have Autism in the Non-Autistic subdirectory, all of which are in the 224 X 224 X 3 jpg format. The unified directory is the second way the data is made accessible. There are two subdirectories in this directory: Autistic & Non-Autistic. It is a collection of files from the train, test, and valid with correct directories that have been combined into a single set. The combined data can then be partitioned into user-defined train, test, and validation sets, and then augmentation technique has been used over our dataset which is used to increase the amount size of the dataset to prevent overfitting and more balance in it.

Several measures which including accuracy, precision, recall, and F-Score are used to measure these pre-trained architectures. MobileNet model with support Vector Machine (SVM) classifier beat the other networks, as demonstrated in Table 4.3, which provides a summary of the results. Furthermore, but using models alone without SVM classifier the MobileNet model is achieved the best results among the other models. Following that, the results of the models employed in this study are listed in order of their accuracy, from highest to lowest for classifying autism. (Densenet121, InceptionV3, Xception, and ResNet-50).

CONCLUSION

As is well-known in artificial intelligence, machine learning, and deep learning, training these networks and models depends on a huge volume of data to train them. There are some techniques used in this thesis which are transfer learning by training deep learning models on our dataset that contains images of Autistic and Non-Autistic. And the augmentation techniques also had been used to add significantly changed copies based on previously current content or freshly made determined by estimating from current data to increase the quantity of data While training a deep learning architecture, it functions as a regularization and tends to prevent over-fitting, address the main problems in training models on the data, filling the gaps, preventing errors and problems of imbalance in the data. After augmenting the dataset, we used our models to train it, and at the last time we used our deep learning models with a machine learning classifier called support vector machine (SVM) on the new version of the dataset (augmented dataset). In this study, five CNN pre-trained models have been used. This CNN pre-trained models are (MobileNet, Densenet121, Inception V3, Xception, and Resnet50). The batch size in our models was 32, the highest number of epochs was 35, the learning rate was 0.0001, with Adam optimizer. As a binary classification, these architectures can identify autistic individuals from other normal people (for example, Autistic and Non-Autistic). By utilizing these strategies and achieving a high level of network evaluation efficiency, early detection of autism is critical so that doctors and patients' families can assist patients adjust to normal life, mix and integrate into the social environment, and enhance their physical and verbal skills, which leads to fewer behavioral problems. In the upcoming we will try other deep learning architectures and improve current architectures and try to collect more datasets to get the best results.

REFERENCES

[1] Crippa, A., et al, (2015). Use of machine learning to identify children with autism and their motor abnormalities. Journal of autism and developmental disorders, 45(7), 2146-2156. https://doi.org/10.1007/s10803-015-2379-8.

[2] Wall, D. P., et al. (2012). Use of artificial intelligence to shorten the behavioral diagnosis of autism. https://doi.org/10.1371/journal.pone.0043855.

[3] Bone, D., et al. (2016). Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion. Journal of Child Psychology and Psychiatry, 57(8), 927-937. https://doi.org/10.1111/jcpp.12559.

[4] Dvornek, N. C., et al. (2018, April). Combining phenotypic and resting-state fMRI data for autism classification with recurrent neural networks. In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (pp. 725-728). IEEE, doi: 10.1109/ISBI.2018.8363676.

[5] Nielsen, J. A., et al. (2013). Multisite functional connectivity MRI classification of autism: ABIDE results. Frontiers in human neuroscience, 7, 599. https://doi.org/10.3389/fnhum.2013.00599.

[6] Aghdam, M. A., et al. (2018). Combination of rs-fMRI and sMRI data to discriminate autism spectrum disorders in young children using deep belief network. J Digit Imaging 31, 895–903. https://doi.org/10.1007/s10278-018-0093-8.

[7] Salman A. D., et al. (2020). ASD classification based on machine learning techniques. journal of xi'an university of architecture & technology, XII(IV). doi:10.37896/jxat12.04/1236.

[8] Beary, M., et al. (2020). Diagnosis of autism in children using facial analysis and deep learning. arXiv preprint arXiv:2008.02890.

[9] Mythili, M. S., et al. (2014). A study on autism spectrum disorders using classification techniques. International Journal of Soft Computing and Engineering, 4(5), 88-91.

[10] Ali. N. A., et al. (2020). Autism spectrum disorder classification on electroencephalogram signal using deep learning algorithm. IAES International Journal of Artificial Intelligence, 9(1), 91. DOI:10.11591/ijai.v9.i1.pp91-99.

[11] Rad, N. M., et al. (2018). Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders. Signal Processing, 144, 180-191. https://doi.org/10.1016/j.sigpro.2017.10.011.

[12] Akter, T., et al. (2021). Improved transfer-learning-based facial recognition framework to detect autistic children at an early stage. brain sciences, 11(6), 734. https://doi.org/10.3390/brainsci11060734.

[13] Raj, S., et al. (2020). Analysis and detection of autism spectrum disorder using machine learning techniques. Procedia Computer Science, 167, 994-1004. https://doi.org/10.1016/j.procs.2020.03.399.

[14] Li, B., et al. (2017). Applying machine learning to identify autistic adults using imitation: An exploratory study. PloS one, 12(8), e0182652. https://doi.org/10.1371/journal.pone.0182652.

[15] Thabtah, F. (2017, May). Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment. In Proceedings of the 1st International Conference on Medical and health Informatics 2017 (pp. 1-6). https://doi.org/10.1145/3107514.3107515.

[16] Stahl, D., et al. (2012). Novel machine learning methods for ERP analysis: a validation from research on infants at risk for autism. Developmental neuropsychology, 37(3), 274-298. https://doi.org/10.1080/87565641.2011.650808.

[17] Wang, J., et al. (2017). Multi-task diagnosis for autism spectrum disorders using multi-modality features: A multi-center study. Human brain mapping, 38(6), 3081-3097. https://doi.org/10.1002/hbm.23575.

[18] Rahman, M. M., et al. (2020). A Review of machine learning methods of feature selection and classification for autism spectrum disorder. Brain sciences, 10(12), 949. doi.org/10.3390/brainsci10120949.

[19] Davarani, M. N., et al. (2017). Identification of autism disorder spectrum based on facial expressions. Emerging Science Journal, 1(2), 97-104. doi: 10.28991/esj-2017-01121.

[20] Sherkatghanad Z., et al. Automated detection of autism spectrum disorder using a convolutional neural network. Front Neurosci. 2020 Jan 14;13:1325. doi: 10.3389/fnins.2019.01325.

[21] Tamilarasi, F. C., et al. (2020, June). convolutional neural network based autism classification. in 2020 5th international conference on communication and electronics systems (ICCES) (pp. 1208-1212). IEEE. doi: 10.1109/ICCES48766.2020.9137905.

[22] Sharif, H., et al. (2019). A novel machine learning based framework for detection of autism spectrum disorder (ASD). arXiv preprint arXiv:1903.11323.

[23] Thabtah, F., et al. (2020). A new machine learning model based on induction of rules for autism detection. Health Informatics Journal, 264–286. https://doi.org/10.1177/1460458218824711.

[24] Liu, W., et al. (2016). Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. Autism Research, 9(8), 888-898. https://doi.org/10.1002/aur.1615.

[25] Kim, M., et al. (2020). deep learning in medical imaging. neurospine, 17(2), 471–472. https://doi.org/10.14245/ns.1938396.198.c1.

[26] Mirsky, Y., et al. (2019). CT-GAN: Malicious tampering of 3D medical imagery using deep learning. In 28th {USENIX} Security Symposium ({USENIX} Security 19) (pp. 461-478).

[27] Rani, P. (2019, November). Emotion detection of autistic children using image processing. In 2019 Fifth International Conference on Image Information Processing (ICIIP) (pp. 532-535). IEEE. doi: 10.1109/ICIIP47207.2019.8985706.

[28] Udayakumar, N. (2016). Facial expression recognition system for autistic children in virtual reality environment. Int. J. Sci. Res. Publ, 6(6), 613-622.

[29] Szegedy, C., et al. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826). DOI: 10.1109/CVPR.2016.308.

[30] Li, X., et al. Classification of breast cancer histopathological images using interleaved DenseNet with SENet (IDSNet). PloS one, 15(5), e0232127. https://doi.org/10.1371/journal.pone.0232127.

[31] Reddy, A. et al. (2019, April). Transfer learning with ResNet-50 for malaria cell-image classification. In 2019 international conference on communication and signal processing (ICCSP) (pp. 0945-0949). IEEE, doi: 10.1109/ICCSP.2019.8697909.

[32] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258). DOI: 10.1109/CVPR.2017.195.

[33] Su, J., et al. (2018, May). Redundancy-reduced mobilenet acceleration on reconfigurable logic for imagenet classification. In International Symposium on Applied Reconfigurable Computing (pp. 16-28). Springer, Cham, DOI: 10.1007/978-3-319-78890-6_2.

[34] Suthaharan S. (2016) Support Vector Machine. In: machine learning models and algorithms for big data classification. integrated series in information systems, vol 36. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7641-3_9.

[35] Gerry. (2020). autistic children data set. kaggle.

[36] Mikołajczyk, A., et al. (2018, May). Data augmentation for improving deep learning in image classification problem. In 2018 international interdisciplinary PhD workshop (IIPhDW) (pp. 117-122). IEEE, doi: 10.1109/IIPHDW.2018.8388338.

[37] Perez, L., et al. (2017). The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621.

[38] Sajja, T. K., et al. (2021). Image classification using regularized convolutional neural network design with dimensionality reduction modules: RCNN–DRM. Journal of Ambient Intelligence and Humanized Computing, 1-12. https://doi.org/10.3390/s21155068.

[39] Bansal, M., et al. (2021). Transfer learning for image classification using VGG19: Caltech-101 image data set. Journal of ambient intelligence and humanized computing, 1–12. Advance online publication. https://doi.org/10.1007/s12652-021-03488-z.

[40] Kumar, G. (2014). Evaluation metrics for intrusion detection systems-a study. Evaluation, 2(11), 11-7.

CURRICULUM VITAE

Abdulazeez Mousa ALMAHMOOD

[redacted]

[redacted] [redacted]
[redacted] [redacted]
[redacted] [redacted]
[redacted] [redacted]
[redacted] [redacted]
[redacted] [redacted]

[redacted]

[redacted] [redacted]
[redacted] [redacted]

[redacted]

[redacted] [redacted]
[redacted] [redacted]

[redacted]

[redacted]

ACADEMIC ACTIVITIES

Paper:

A. A. Mousa and F. Özyurt, "Autism detection from facial images using deep learning methods," 2021 International Conference on Advanced Engineering, Technology and Applications (ICAETA), 2021, pp. 24-28.

B. A. Mousa, M. Karabatak and T. Mustafa, "Database Security Threats and Challenges," 2020 8th International Symposium on Digital Forensics and Security (ISDFS), 2020, pp. 1-5, doi: 10.1109/ISDFS49300.2020.9116436.