A deep learning approach for prediction of ADHD using brain structure

by

Shafin Ahad Siam
23341080
Durjoy Datta
19301208
S.M.Kawsar Hossain
19301221
MD.Nishat Ahmed Nobel
19301133
Ashik Ahamed
19301123

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
School of Data and Sciences
Brac University
September 2023

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

SHAFIN AHAD SIAM

Shafin Ahad Siam
23341080

Durjoy Datta

Durjoy Datta
19301208

Kawsar

S.M.Kawsar Hossain
19301221

Nishat Ahmed

MD.Nishat Ahmed Nobel
19301133

Ashik

Ashik Ahamed
19301123

Approval

The thesis/project titled "A Machine Learning Analysis for Predicting Children with ADHD Using Behavioral Activity" submitted by

1. Shafin Ahad Siam(23341080)
2. Durjoy Datta(19301208)
3. S.M.Kawsar Hossain(19301221)
4. MD.Nishat Ahmed Nobel(19301133)
5. Ashik Ahamed(19301123)

Of Summer, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on September, 2023.

Supervisor(s)

Supervisor:
(Member)

Dr. Farig Yousuf Sadeque
Assistant Professor
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

Mr. Md. Tanzim Reza
Lecturer
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Attention deficit hyperactivity disorder (ADHD) is a complex condition affecting the brain's neurodevelopmental processes. Prompt treatment and accurate diagnosis can alter neural connections and improve symptoms. This article focuses on the classification of MRI scans of individuals with ADHD and those without the disorder using deep learning algorithms. Pre-trained models like VGG-16, RegNet-50, and DenseNet-121 were used, along with non-pretrained models like CNN and convolutional LSTM. VGG16 is known for its intricate architecture, while CNNs have undergone significant advancements, resulting in improved classification accuracy. The convolutional LSTM model, a novel integration of CNNs and LSTM networks, was used to predict ADHD based on anatomical data from the brain. The results showed that only the VGG-16, CNN, and convolutional LSTM models exhibited superior accuracy. Ensemble learning was used to create an ensemble model, with convolutional LSTM having 93% accuracy, CNN having 95% accuracy, and ensemble learning having 97% accuracy. Overall, ensemble learning had the highest accuracy, indicating the need for a model that can detect ADHD with better accuracy.

Keywords: Deep learning, Convolutional Neural Network, Attention deficit hyperactivity disorder, Ensemble learning, diagnosis, treatment, artificial intelligence, magnetic resonance imaging (MRI), VGG-16, RegNet-50, DenseNet-121, convolutional LSTM, convolutional neural networks (CNNs), evaluate, accuracy.

Acknowledgement

Firstly, all praise to the almighty for whom our thesis have been completed without any major interruption.
Secondly, to our advisor Dr. Farig Yousuf Sadeque sir, co-advisor Mr. Md. Tanzim Reza sir for their kind support and advice in our work. They helped us whenever we needed help.
Thirdly, The whole department of computer science and technology of BRAC University for giving advices in certain problems through research assistants or honorable faculties, which gave us immense confidence to finish our thesis
And finally to our parents, without their throughout sup-port it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

ADHD Attention deficit hyperactivity disorder

CNN Convolutional Neural Network

DenseNet Densely-connected Convolutional Network

EDA Exploratory Data analysis

fMRI functional Magnetic Resonance Imaging

LSTM long short-term memory

ResNet Residual Neural Network

ROC Receiver Operating Characteristic

sMRI Structural Magnetic Resonance Imaging

VGG Visual Geometry Group

Chapter 1

Introduction

1.1 Overview

Attention deficit hyperactivity disorder is one of the most typical neurodevelopmental behavioral problems in kids (ADHD).[32] Impulsivity, hyperactivity, and inattentiveness are common traits in children with ADHD. To stop childhood ADHD, researchers are attempting to identify the risk factors. An association between genetic factors and ADHD was found, as was the significance of these factors. Impulsivity and issues with social engagement are two behavioral characteristics that both ASD and ADHD share. To narrow down the differential diagnosis, clinicians must contend with behavioral similarities. A kid with predominant ADHD may struggle socially, but this may be due to impulsive interruptions or an inability to pay attention to what is being said rather than a basic misunderstanding of social signals, which is more typically evident in children with autism.[9] This is not meant to suggest that ADHD and autism cannot coexist. The Diagnostic and Statistical Manual (DSM-V), which had previously prohibited a dual-diagnosis under DSM-IV, changed its diagnostic criteria to clinically legitimize the autism-ADHD comorbidity. This is due to the fact that it has acknowledged how frequently ASD and ADHD symptoms co-occur.[7] Increasing evidence suggests that individuals diagnosed with ADHD exhibit distinct variations in brain structure and function. Neuroimaging studies have revealed that changing in neuronal connections of some brain regions and adjustments in brain volume. These neuroanatomical differences are associated with cognitive function, motivation regulation, and attention in individuals with ADHD. The brain's reward system, primarily regulated by dopamine neurotransmission, is also affected in ADHD. Notably, individuals with ADHD often exhibit abnormally low presynaptic dopamine storage in the prefrontal cortex, leading to impaired attention, cognitive processes, and active memory. Due to this deficiency syndrome, people with ADHD are more likely to engage in violent behavior, drunkenness, drug addiction, and other behaviors that boost dopamine production. There is a higher risk of SUD for individuals with comorbid conduct disorder, with a rate four times that of population. Emerging evidence suggests that ADHD-related neuroanatomical and functional alterations could not last forever. Quick recognition and based on research treatment of ADHD symptoms can result in neurons and function that are similar to those of neurotypical people.[7] Adults who have previously overcome from ADHD have been found to have normalized brain connections as opposed to those who still experience symptoms. Successful management of ADHD can significantly improve quality of life and facilitate better integration into society. Thus, early identification of ADHD and evidence-based management strategies are crucial for optimizing long-term outcomes. Since some people with ADHD may go undiagnosed or be underdiagnosed by existing clinical tests, adjunctive data sources may also be helpful in identifying the complete range of ADHD sufferers. Numerous research have looked into the possibility of using neuroimaging modalities including Magnetic Resonance Imaging (MRI), physiological signals like electroencephalograms (EEG) and electro-cardiograms (ECG), accelerometers, and game simulators. These studies attempt to lessen the workload of clinical diagnosticians by suggesting AI advances, especially machine learning (ML) and deep learning (DL), for quicker and more affordable ADHD diagnosis. Machine learning (ML) methods have become a potential new tool for the identification and diagnosis of a number of illnesses, including ADHD. Large-scale data analysis, the discovery of intricate patterns, and precise prediction based on learned patterns are all capabilities of machine learning algorithms. Researchers and physicians want to create fast, objective tools for the early identification of ADHD by utilizing ML algorithms, which will also help to improve patient outcomes and the diagnostic process. To illustrate the idea, we trained a machine learning model which is CNN sequential to anticipate the diagnosis of ”ADHD” in a clinical sample that required assistance. To train this model, the dataset has been converted to images by splitting into dimensions for the purpose of testing.[7]

1.2 Research Objectives

By developing a machine learning detection approach for diagnosing Attention Deficit/Hyperactivity Disorder, this project aims to establish a platform where patients may get therapeutic assistance and mental health care through ongoing monitoring and engagement. One of the main objectives of the study is to have a thorough understanding of ADHD and its impacts

1. To know how ADHD is identified,
2. To develop a machine learning model for identifying and helping those with ADHD,
3. To make life simpler for persons with ADHD and to establish a model that would help them to identify ADHD,
4. To submit our research to prominent journals or conferences for publication.

1.3 Research Problem

Over 75 percent of young children's ADHD risk is attributed to genetic factors. [32]. To lower the number of children who have the illness, researchers are attempting to identify the risk factors connected to ADHD. LR may be applied as an FS technique in this inquiry to pinpoint the most significant risk factors for kids with ADHD.

Age, sex, asthma, allergies, depression, anxiety, use of alcohol, insurance status, race, family structure, and socioeconomic level were all risk factors for ADHD in children. The following experimental aspects may be predicted by ML-based classifiers based on the risk factors for ADHD: baseline and demographic parameters, bal-

anced dataset construction, and selection of the salient significant risk variables.[32] Many challenges confront those who have ADHD. Intermingling colors is one of them. As a behavioral indicator for RST, it is applicable. Using NIRS devices, the RST is performed while the subject's brain activity is monitored based on oxy-Hb levels.It is possible to train the data using the knowledge this system has acquired.[27]
Because adult ADHD is a combination of numerous chronic diseases, it might be challenging to comprehend that person's behavior. So, it can be observed in some tasks. While doing MRI tests, their action on task and rapid changes in attention can affect MRI. This MRI data set can be transformed into images. There will be different changes in these pictures. Using a CNN model, this process will get more accuracy and be less time-consuming.
It takes more time and money to diagnose attention deficit hyperactivity disorder, which is a considerably more complicated condition. Additionally, crucial information is needed for differential diagnosis and behavior-based ADHD detection. Thus, it makes good therapy difficult to start. Currently, automated assay makes it less complicated and takes less time. The complexity of the diagnostic procedure can be decreased with the automated test. Additionally, the clinical sample may be used to train the data so that it can distinguish between the patient's various behavioral activities.[33]
Clinical samples are one of the finest sources of training data for the detection of ADHD. Clinical samples may be classified into aspects including demographic factors, symptom ratings, neuropsychological assessments, etc. [33]. It is simple to teach our staff to divide the population based on groups and save a lot of time by using neuropsychological measurements to divide the ADHD and non-ADHD groups based on IQ levels.

There are 15 million persons with ADHD worldwide [31] . The percentage of youngsters in this group is 7.2%. Since ADHD has major repercussions and numerous linkages to other illnesses, early symptoms of the disorder are challenging to identify. In order to lessen the level of difficulty, machine learning techniques can be used to this identification problem. A excellent method for forecasting a person's behavioral changes is to use neurobiological measurements.
Impulsivity is the most crucial factor. Impulsivity is the tendency to act without first considering. Therefore, it may be a significant issue in our society. The likelihood that someone with impulsivity also has ADHD is quite high. Our methodology can be really helpful in resolving this issue. People should engage in mindful practice to address this issue. The author of "The Mindfulness Prescription for Adult ADHD," board-certified psychiatrist Lidia Zylowska, advises "bring attention to the present moment and observe what is happening without criticizing it."

Chapter 2

Literature Review

As Google scholar is considered one of the best sites for collecting information, especially for Thesis research so we used Google scholar to search for different research papers. We used different keywords to search like ”ADHD” ”Machine learning” etc. We researched a good number of research papers and here we will review some good research papers similar to our topic.
To lessen the number of kids with ADHD, Moniruzzaman et al. sought to identify the risk factors. Genetic factors account for over 75% of the risk of ADHD in younger children. [32]. Risk factors include brain damage, alcohol and cigarette use during pregnancy, and early birth. Based on a continuous performance test, Slobodin et al. also identified children with ADHD (CPT) They chose 458 kids between the ages of 6 and 12.The best accuracy, sensitivity, and specificity (88.0 percent, 89.0 percent, and 84.0 percent, respectively) were shown to be provided by the ML-based classifiers (MOXO). For descriptive analysis, we utilized Stata 14.10, and for ML-based analysis, we used Python 3.9 and Scikit-Learn 1.2. In statistics and machine learning, feature selection is sometimes referred to a variable selection (ML). FS is a method for determining which characteristics are the most instructive so that ML-based algorithms may perform better. In this study, the most important risk variables for children with ADHD were extracted using the LR as an FS approach. 90% of the dataset was utilized for training, while the remaining 10% was used for testing. In order to predict ADHD in children between the ages of 3 and 17, we used a feature selection approach and eight ML-based classifiers in this study. Prior to establishing the class classification, 11.4 percent of people in the population had ADHD. According to research findings, anxiety and sadness affected 37.6% and 41.9% of children with ADHD, respectively. If a kid was raised in a two-parent household, their likelihood of being given an ADHD diagnosis was much reduced. Current behavioral symptom overlaps make the diagnosis procedure for ADHD time-consuming and challenging. Given the high prevalence of ADHD, a technique that can quickly and accurately assess the risk of ADHD is required.[32]

Ruth Gwernan-Jones et al. claim that the Diagnostic and Statistical Manual of Mental Disorders (DSM) now refers to ADHD as a neurodevelopmental condition characterized by levels of impulsivity, hyperactivity, and inattention that are out of proportion to one’s age. [8] This study, which examines non-pharmacological ADHD therapies in schools, is one of many systematic ones. The search approach consisted of three parts: words or phrases that alluded to ADHD, a school environment, and a special qualitative research filter. The review's major theme was how mothers were silenced—both by other people and by themselves. Rising opposition, disappointed expectations, and parent-teacher conflict as the norm were some other themes. Each will be discussed in more detail below. On a web forum for moms of kids with ADHD and/or learning issues, it was found that conflicts with schools were the most often mentioned topic. Hibbitts referred to five instructors as her son's "life lines" throughout his academic career: a SENCo, two headteachers, a classroom teacher, and a school counselor. Teachers' impressions of themselves as professionals were appreciated more than mothers' assessments of themselves as "amateur parents." The author's family experienced cultural dissonance, which led to unpleasant and false perceptions about her children and family. She claims that similar indicators are probably present in all societies. It was acknowledged that the need for an ADHD diagnosis or concern about symptoms was a byproduct of the educational context. The options to resistance were not equitable; those who took the greatest initiative (homeschooling, sitting on school boards, or paying for private tuition) had more access to resources in terms of money, knowledge, and social connections. Parents have reported that certain instructors, despite their politeness, have a tendency to dismiss the information on ADHD that parents intended to share with them. Some mothers thought that teachers felt that knowing about their children's upbringing diminished their position as professionals. Malacrida compared giving instructors access to knowledge to what Michel Foucault referred to as "truth games" In Malacrida's study, mothers occasionally began to refuse to collaborate after making an attempt to appear polite and obedient, entering a more powerful phase of resistance. This took the form of making a public statement by joining advocacy groups, choosing a different school, or choosing to homeschool their kids.[8]

Researchers who study attention-deficit/hyperactivity disorder (ADHD) have long been interested in Sarah Durston et al.'s expressed regulation of action and cognition[19]. This skill, which involves suppressing improper behaviors in favor of appropriate ones, is frequently referred to as cognitive control in ADHD studies. It has also been proposed that this disease is primarily caused by issues with cognitive control. Only 30–50% of kids with ADHD do poorly on assessments of this skill. A considerable heritable component contributes to the prevalence of the condition ADHD. Because a condition's associated phenotype is linked to a biological pathway, endophenotype approaches by their very nature need a theoretical approach to a condition. In this work, we examine the research on endophenotypes of ADHD based on neuroimaging measurements and evaluate the usefulness of cognitive control as such a measure. At least in terms of cognitive control, ADHD is a highly heritable illness with a clear neurological underpinning. For monozygotic twins, concordance rates range from 50% to 80%, whereas for dizygotic twins, they are closer to 30%. (Bradley and Golden, 2001, Thapar et al., 1999). On the heredity of brain activity and structure in ADHD, there is currently a dearth of information.[19]

ADHD is a neurodevelopmental illness that frequently lasts into adulthood and manifests in childhood, according to Nicola Wright et al [3]. Estimated prevalence rates range from 3 to 9 percent in the school-age population of the United Kingdom (NICE, 2008). If left untreated, ADHD can lead to difficulties in a variety of areas, including academic achievement and parental efficiency. Additionally, there is a higher chance of behavior and personality problems, drug abuse, and poor social skills. It has been determined that the issue known as "access to care" is widespread, varied, and complicated, including components from sociology, psychology, management, economics, and epidemiology. There are extensive supplemental literatures on topics like healthcare quality, priority setting, and patient satisfaction, and its definition has not been regularly used (Dixon-Woods et al., 2006). For this review, Gulliford et alconcept .'s of access to care was used (2001). We have chosen to adopt a wide socio-ecological perspective in order to illuminate the social, cultural, and interpersonal interactions that shape how individuals behave (Susser Susser, 1996). We also wanted to highlight the areas where further research is needed and where there is evidence that may be used by both individual healthcare providers and larger healthcare institutions. Whether issues were conceptualized as "emotional challenges" or ADHD, there was a higher inclination for referral. Parents of children with ADHD saw their child's behavior in a scenario as less severe than parents of children with less signs of ADHD, according to research by Maniadaki et al. (2006). According to Ohan and Visser (2009), this belief system mediated the link between gender and service seeking and affected parents' and instructors' judgment. A study of the literature looked at how ADHD affects children's and adolescents' access to services. It was discovered that a positive ADHD status had the most impact on whether someone was recognized. Gender and ethnicity of the kid seem to have an impact on how they obtain services. Comorbidity was shown to be a significant predictor of access to care, and cultural disparities in parental thresholds for labeling behavior as problematic were detected (Gidwani et al., 2006). School-based therapies may be beneficial given the effect of ADHD on functioning in educational environments. For a number of pediatric illnesses, obstacles and facilitators to care have been identified at the individual and organizational levels. The question of whether ADHD exists or not has a broader political component. Practitioners of mental health care need to be aware of how teenagers view the stigma associated with ADHD as a barrier to participation in the screening process. In order to guide efforts to enhance fairness in access to care, clinicians who deal with children and adolescents will benefit from a greater knowledge of these hurdles.[3]

Autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD) are two of the most prevalent conditions in children, according to M. Duda et al. [10]. Significant behavioral parallels occur between ASD and ADHD, including impulsivity and trouble socializing. ADHD and autism commonly co-occur as symptoms, according to the most current Diagnostic and Statistical Manual (DSM-V). We created accurate classification techniques that successfully discriminate ASD from non-ASD. We merged the full item-level score sheets from the SRS for 150 participants with ADHD and 2775 individuals with autism. The ADHD individuals did not have any autism diagnoses, while the ASD patients did not have any documented ADHD comorbidity. Six machine learning algorithms were developed and evaluated using data on autism spectrum disorders and attention deficit hyperactivity disorder (ADH) (ASD). Utilizing the linear kernel, several Logistics Regression models, and Logistic Regression with l1 regularization, the classification (SVC) model was created. When additional characteristics were added to the model, the majority of algorithms' performance gradually decreased. Using a small subset of behaviors collected from the SRS, we looked at whether it would be feasible to distinguish between ASD and ADHD. We developed classifiers using machine learning to separate these clinical groups using just a few criteria. Using the various machine-learning methods we employed, we were able to create models that were optimized on the observed data, which resulted in the development of new classification criteria. The outcomes showed that the presented classification task was not a good fit for the two tree-based algorithms Decision Tree and Random Forest. Four of our six algorithms, which scored well, required only five behaviors (AUC40.96). As a consequence, 492 fewer behaviors were analyzed using the standard SRS. Although the SRS's suggested age range is 4 to 18, younger children may utilize it owing to the tool's very adaptable behaviors. Given that the average age of diagnosis for ASD and ADHD is currently 4.5 years and 7, respectively, there is a major motivation to focus attention on developing speedier and more reliable approaches for risk detection. In order to improve throughput and backlog management and improve clinic patient flow, a mobile screening tool may be beneficial.[10]

Ilke Oztekin et al. concentrated on the characteristics of ADHD, a developmental disorder marked by impulsive, hyperactive, and inattentive behavior [31]. Early ADHD diagnosis and treatment are deemed essential due to the serious effects. This study evaluated the predictive accuracy of executive function (EF) assessments in 162 children between the ages of 4 and 7. Overfitting is a serious impediment to the creation of a valid prediction model that may offer results that are transferable to a clinical diagnosis. The AHEAD research was among the first to scan kids with ADHD when they were as young as 4 to 7 years old. Imaging research has not focused much on the neurology of ADHD in this young age range. The AHEAD (ADHD Heterogeneity of Executive Function and Emotion Regulation Across Development) research aims to define the range of traits that have been associated with ADHD in early children. Among the first to screen kids as early as 4 to 7 years old for ADHD. Diagnosing ADHD early on is essential for treatment to be successful. Executive function processes and ADHD have previously been connected in research (Barkley, 1997; Sergeant,2000; Sonuga-Barke, 2002). Given the implied role of executive function in thecondition, we hypothesized that our models would successfully predict the diagnostic categories for ADHD. Children and the individuals who cared for them were sought out via brochures, radio and print commercials, open houses, and parent seminars. The concerns with disruptive behavior and ADHD were the focus of this study. Participants needed to have an estimated IQ of at least 70, be able toattend an 8-week summer treatment program before the start of the next academic year, and have attended school the previous academic year (ADHD groups only). Given the strong association between ADHD and executive function demonstrated in other studies, our features serve as parent/teacher evaluations of executive function. One of these indicators is how well you do on the three tasks that test your executive function (Flanker task, Dimensional Card Sorting task, HTKS task).[31]

ADHD is a significant behavioral trait and is crucial for clinicians to fully comprehend, according to Aki Nikolaidis et al[31]. Numerous genes have been examined in studies on the genetic genesis of the illness. One of the most dependable possible polymorphisms in the quest for a genetic link to ADHD is the 48-base pair repeat in the 7-repeat allele of the DRD4 gene. Although several pertinent meta-analyses have been published, they have, to our knowledge, all been Caucasian or Asian subject-specific. From more than 3200 publications collected from PubMed and Google Scholar, 600 abstracts on subjects pertaining to ADHD were kept. The publications we chose that had a direct connection to the 7 repeat DRD4 allele helped us to focus our search (exon III VNTR).Due to the very low percentage of Asians who have the 7R mutation, we decided to consider papers with 95 percent or greater racial homogeneity (Chang et al., 1996) There were also three levels of ADHD recruiting: low, medium, and high. Subjects who were diagnosed with ADHD based just on an interview with the child's teacher or parent were referred to as having low rigor. High rigor meant that at least one psychiatrist had given the ADHD individuals an official diagnosis. We used Comprehensive Meta-analysis to analyze this data utilizing software capable of calculating a broad variety of meta analytical statistics. We saw a fall in i2 to 66.7 after eliminating Kim et al., 2005 in addition to a substantial reduction in the association between the 2R allele and ADHD citer7. Since both Caucasian and South American groups have a comparable link between the 7R and ADHD. High OR 1.92 (1.44-2.55, Z = 4.49) was shown by low confidence grouping. For Caucasians and persons from the Middle East, there were highly different levels of correlation between ADHD and the DRD4 7-repeat allele. Strength of the connection was unaffected significantly by the accuracy of the ADHD diagnosis. By classifying individuals based on their ancestry, future research in the field of mental genetics might readily add cultural uniqueness to its findings. The variations in connections between DRD4 and ADHD between persons from Europe and the Middle East might potentially be explained by variations in the diagnosis of ADHD. Studies have revealed that race and genotype may have an impact on how children respond to and understand parenting behavior. The difference between racial groups may also be influenced by socioeconomic level, which has been shown to limit externalizing behavior. Positive selection has most certainly been working on the 7R allele for a very long time. The difference in ADHD DRD4 7R relationship between Middle Eastern and European Caucasians may be influenced by both genetic and cultural factors. [31]

The most common mental health issue affecting young children, according to Anette Primdal Kvist et al., is ADHD [6]. 3.7% of children in school-age have the disease, which is characterized by attention difficulties, hyperactivity, and impulsivity. The goal of this research is to investigate how having an ADHD child affects both the job market and the breakdown of parent-child relationships. The existence of an ADHD kid affects both the chance that a relationship (marriage or cohabitation) will terminate and the availability of both parents' labor. Both among papers particularly addressing ADHD and among those investigating child impairment more generally, our work is the first to use data from a statewide administrative panel. Long before ADHD is ever diagnosed, the consequences for parents may begin to take shape. Males experience the disease more often than females, with a gender ratio of 6:1 to 9:1 in clinical settings. Maternal smoking, pregnant alcohol use, challenging pregnancies and deliveries, and low birth weight are all prenatal risk factors. Even while the labor supply and the stability of parent-child relationships may be related, we nonetheless regard them as two separate outcomes. In order to calculate the risk that the couple has broken up, we estimate probit models for each of the first ten years after childbirth. Maternal psychopathology and other unobserved confounding factors cannot be adequately defined in register-based data. In a system where a nonspecialist rather than a specialist makes the diagnosis, concern concerning misdiagnosis is more significant (such as in the United States). We expect non-specialists to mistake stress symptoms in children for ADHD symptoms more frequently. According to previous research, 89 percent of individuals found in the Danish Psychiatric Central Register would also be given a full diagnosis of ADHD (DSM-IV).Research from the past has shown the reliability of the ADHD diagnoses included in the Danish registry. Due to the fact that we exclusively research people who receive a diagnosis at general hospitals, There was no basis to question whether the diagnoses documented in our data were accurate (rather than private clinics). Without ever obtaining a diagnosis, some children may suffer with ADHD.[6]

Diagnosing and treating attention deficit hyperactivity disorder (ADHD), as Pavol Mikolas et al. have hypothesized, is arduous, time-consuming, and costly [27]. The diagnostic process has to be simplified, clarified, and made shorter. This could be possible if machine learning techniques are used. Problems with attentional processes are explicitly referred to as "ADHD" in this context. Issues with working memory and processing speed, as well as having a lower overall IQ15, have all been proposed as markers that distinguish persons with ADHD from normally developing controls. In this proof-of-concept work, we aimed to teach an AI model to anticipate the diagnosis of "ADHD" in a clinical sample that was seeking treatment. Attention deficit hyperactivity disorder (ADHD), attention deficit disorder without hyperactivity, and hyperkinetic conduct disorder (F90.0) diagnoses were classified as "ADHD" patients (F98.80). Males, younger people, and those with mental comorbidities were all significantly more likely to be diagnosed with ADHD (chi2 = 6.871, p + 1.009). The data sets were removed if neurological, genetic, endocrine, or other significant, known medical conditions were discovered. When performing the study, the Declaration of Helsinki was observed. To divide the subjects into groups with and without ADHD, we employed a linear SVM classifier. The 19 most precise traits from the attention, intelligence, and symptom assessment categories were combined to provide the most efficient performance. There was no age or gender matching between individuals with ADHD and those without it in this population-based research since it was population-based. Using datasets without any missing data might boost precision even further. Many people asked for aid.Other conditions in our sample's participants were also found, some of which could exhibit symptoms similar to those of ADHD (such as attention deficits in depression, increased activity in tic disorders, etc.). This may be as a result of the two categories not being distinguished (ADHD vs. "something else"). The average IQ score did not come in as one of the most reliable indicators. The processing speed subscale placed top in a classification of all IQ-related factors. This may be because earlier studies compared individuals with ADHD to healthy controls and discovered that this aspect of neuropsychological processing was significant. We were unable to include more extensive clinical markers, such as the CBCL, as prospective features because of the high amount of missing values. By employing bigger samples with more full data on all clinically important criteria, classification performance should be significantly improved. [27]

Sigurd Ziegler et al. tried to determine whether attention deficit hyperactivity disorder (ADHD) is a prevalent neurodevelopmental issue that has negative effects on scholastic performance, professional success, and health. Decision-making (DM) and reinforcement learning (RL) are modified in this case.[12] Understanding DM and RL is aided by computational modelling since it brings together behavioural and neurobiological data. Predictions for the influence of ADHD on DM and RL are detailed in this review of neurobiological theories of ADHD, using the drift-diffusion model of DM (DDM) and a simple RL model. The empirical evidence supports the theoretical predictions of a decreased DDM drift rate in ADHD, whereas the results of border separation are less certain. outline areas for further theoretical development in the field of ADHD. The foundations of RL and DM have been untangled through the use of computational modelling. It can be more sensitive than conventional assessments, but only if clear hypotheses are made about the links between the operations of the mind and/or the brain. This link between behavioural and neurobiological findings is crucial for our comprehension of neurodevelopmental disorders.Although most neurobiologically detailed theories concur on the involvement of the brain's catecholamine systems in ADHD, their hypotheses regarding the specific neurobiological mechanisms underlying ADHD diverge, indicating that a consensus on the cognitive and neurobiological basis of ADHD has not yet emerged. Using two well-known computational models, this study investigates the underlying pathogenic mechanisms proposed by the most neurobiologically specific ADHD hypotheses. It evaluates the plausibility of their primary neurobiological assumptions and predictions for DDM and PERL characteristics for four well-known hypotheses of ADHD. The predictions are subsequently contrasted both within and between theories, as well as with recent actual findings from investigations employing suitable computer models. Experiments are advocated by the authors as a means of validating hypotheses. The neurobiology underlying ADHD, rational choice, and reinforcement learning Models provide mechanical explanations of catecholamine activity in RL and DM; consequently, there is a growing interest in employing computational models to bridge human physiology and behaviour. This article examines the pathogenic pathways provided or suggested by the most influential computational models of DM and RL in order to gain a better understanding of ADHD. The findings lend credence to the notion that DDM and PERL modelling may be used to characterise the neurobiological assumptions underlying ADHD and its symptoms. [12]

Ms. Sumathi M.R. et al. told The quality of life can be greatly enhanced through the early detection of mental health issues.[11] Medical records are analysed by machine learning methods so that five common mental disorders can be identified and treated. Three classifiers were shown to be more effective than the whole set of attributes in predicting class membership. Researchers in artificial intelligence have created machine learning algorithms for making correct diagnoses of mental health issues. The top three methods for helping mental health practitioners diagnose five common mental health issues in kids have been discovered in this study. DTREE, INTERNIST/AUTOSCID, MILP, BPNN, ANFIS, MCDA, Basavappa S.R., et al., comparing different classification systems, have all been utilised in the past to aid in the diagnosis of mental health issues. Diabetic, Parkinson's, Alzheimer's, and mild cognitive impairment diagnostic accuracy varied across data mining techniques. When compared to other classifiers, the Bayesian Network (BN) Decision Model excelled even when only a few polymorphisms were present. A quartet of feature selection algorithms succeeded in achieving the best accuracy. The accuracy of identifying mental health illnesses is being improved by using a variety of machine learning techniques. In this study, a number of different machine learning and classification strategies, such as AODEsr, MLP, RBF Network, IB1, KStar, Multi-Class Classifier (MCC), FT, and LADTree, were put to the test to determine how well they could forecast primary mental health disorders. The study's diagnostic methods included back propagation, RBF networks, IB1, K*, multiclass classifiers, FT functional trees, LAD trees, and logical analysis of data trees for the five most frequent mental health disorders. The data set collected from a clinical psychologist was analysed by hand, and the results led to the discovery of 25 features. Through the use of the Best First Search method, extraneous characteristics are removed from the dataset before further analysis. The identified characteristics are as follows: age, sustained attention, family history, Mani episode test score, relationship formation, anxiety disorder, major depressive episode, focus, CDI score, restlessness, PDD score, seizures, and the severity of autism are all measured here. Classification algorithms' efficacy can be compared using several metrics, including the Kappa statistic, accuracy, and ROC area, all of which are available on the WEKA platform. Accuracy indicates how well a classifier performs on a collection of test examples, while Kappa represents how well the classifier agrees with the real class. When it comes to identifying children with basic mental health disorders, the multilayer perceptron, the multiclass classifier, and the LAD tree are three examples of classifiers that are superior to others in terms of accuracy.[11]

ADHD is an inhibitory control disorder, and problems with working memory and short-term memory are characteristic of disruptive behaviour disorder, as found by Jack Stevens et al. Russell Barkley published a model to describe and predict the numerous challenges faced by children with ADHD. [2] The stop-signal task is a potential instrument for assessing inhibitory control in children with ADHD due to its simplicity, rapid administration, and capacity to rule out aversions to delays as a cause of poor performance. Because it does not require learning difficulties and can be made simpler or more difficult, the stop-signal paradigm may be useful for assessing inhibitory control in children with ADHD in reward-versus-no-reward settings. The effect of behavioural inhibition on the working memory of students diagnosed with attention deficit hyperactivity disorder (ADHD) was investigated. In terms of their abilities to control their conduct, their motivation, and their working memory, children with ADHD had a higher opinion of themselves than children without ADHD. The social position scores of the parents of children with ADHD were higher than those of the control group, while the estimated IQ scores of the children themselves were significantly lower than those of the control group. ADHD = attention deficit hyperactivity disorder; control group families had median annual incomes between $30,000 and $39,000; fathers had higher ages and worked more hours; and N = 76. The Conners Parent and Teacher Rating Scales (Revised) and a demographic questionnaire were used as measures. High criteria for concordance and agreement. Based on the parents' levels of education and professions, a Hollingshead social position score was generated. Behavioural inhibition was evaluated using the stop-signal task and the Kaufman Brief Intelligence Test. The timing of responses to a "stop" signal served as an evaluation of inhibitory control. Children with and without ADHD completed two tests with exceptional split-half reliability, one with and one without a monetary reward. Processing and storing stimuli, as well as short-term memory, were used as tests of working memory. The children were tested on their ability to recall numbers from two different lists. The number of numbers accurately recalled in any order had greater split-half reliability than the number of trials completed. The tasks' complexity and the children's performance were evaluated based on their ratings. When comparing the ADHD group to a control group, hyperactivity, impulsivity, inattention, anxiety, perfectionism, somatic complaints, and social problems were all evaluated higher in the ADHD group. There were tests of age, memory, IQ, parental social status, and reinforcement, but no reinforcement for inhibiting behaviour. The results showed that children with ADHD had far slower reaction times in response to a stop signal compared to the control group. There were no results that indicated a connection between group membership and other variables. There was no difference in how well people did on the other metrics of the stop-signal challenge. During the experimental manipulation, children with ADHD did much better when they were rewarded. There were substantial differences in terms of age, group, order of events, reinforcement, and memory tasks. Socioeconomic background, age, and group engagement were all areas where the ADHD and control groups diverged significantly from one another. Since the ADHD group remembered fewer digits correctly and took longer to process coloured stimuli, stronger inhibitory control was connected to better working memory. [2]

According to research by Sharath Chandra Guntuku et al., adults are frequently misdiagnosed with attention deficit hyperactivity disorder (ADHD), a disease characterised by impulsivity, hyperactivity, and inattention.[16] Using social media accounts is an innovative approach to gathering naturalistic data on the experiences and public expressions of people with ADHD.Researchers can gather potentially important insights for researchers and doctors by studying the social media behaviour of users with ADHD, and patients with ADHD can receive feedback through this platform. In order to better understand the behavioural symptoms and comorbidities associated with ADHD, this study analysed the vocabulary of Twitter users who had been diagnosed with the disorder. Open-vocabulary techniques are becoming increasingly popular for analysing social media, and data-driven approaches propose hypotheses for verification or rejection. Anxiety, sadness, and post-traumatic stress disorder were among the illnesses that might be predicted using open vocabulary techniques, which were put to use in the CLPsych workshop. In order to decipher the themes that people with ADHD discuss in a massive data collection, this research employs a differential language analysis based on an open vocabulary. Excluding retweets and tweets written in languages other than English, we were able to collect 1.3 million public posts from all individuals with ADHD. Depression, anxiety, and bipolar disorder were common mental health diagnoses among the eleven individuals, who all had similar medical histories.Words, phrases, subjects, and LIWC traits were retrieved based on their relative frequency across all tweets. To understand how people with ADHD think, we employed closed-vocabulary analysis (LIWC) and differential language analysis (DLA). Word tokens across several LIWC classes were found to be significantly correlated with ADHD symptoms. There is a strong correlation between ADHD and the use of words implying doubt and hedging behaviour on social media, which may indicate concerns or an apprehension of the unknown. Words expressing anxiety are substantially connected with ADHD on social media, suggesting that two conditions often occur together. The use of a wide range of terms is intended to facilitate communication. People with ADHD are more inclined to cuss, gossip, and discuss Pokemon Go amongst themselves. Their messages also feature elements of their personalities, a sense of time, and drug use. These traits may indicate problems with self-control. This study analysed the online behaviour of people who self-reported a diagnosis of ADHD on Twitter and found significant variations in self-efficacy, emotional dysregulation, negation, self-criticism, substance usage, and displays of mental, physical, and emotional weariness. Social media posts provide a transparent look into one's state of mind, and this study found that adults with ADHD are more likely to participate in negative or maladaptive thought patterns than members of control groups. These realisations could guide the development of hypotheses for studying the various ADHD presentations. Due to the disorder's possible impact on adults' sense of competence, those diagnosed with ADHD are more likely to use a social networking platform that allows for instant and rapid responses to their posts. They have difficulty expressing themselves, which is reflected in their tendency to dwell on the past rather than the present or future. Out-of-sample accuracy for predicting ADHD based on social media language was found to be 78.5%. Clinical judgement and social media data could be combined in future research to better screen for and diagnose ADHD. In order to gain some understanding of the perspectives and feelings of individuals with ADHD, this study used a linguistic analysis of online posts.[16]

According to Andreas Mueller et al., ADHD is a Inattention, hyperactivity, and impulsivity are hallmarks of attention deficit hyperactivity disorder (ADHD), a neurobehavioral condition.[5] There are repercussions in terms of money, an emotional toll on families, and professional and educational opportunities lost as a result. In this study, we analyse how different ERP components define the neurocognitive information processing system. Researchers have used ERPs, electrical representations of brain processes, to examine sensory and cognitive-processing anomalies in ADHD. We used independent component analysis (ICA) to separate the EEG signals. The development of non-linear and non-parametric classifiers in the field of electroencephalography has allowed for the objective detection and diagnosis of anomalies, as well as the evaluation of treatments. ICA is a computational approach for deconvolving muddled signals into their component parts. In this work, researchers looked into the feasibility of utilising a non-linear support vector machine classifier to categorise adult ADHD patients based on their ERP data. Participants who already had a diagnosis of ADHD were placed in this group. A local ethics commission reviewed and approved the written informed consent that all participants provided. A Mitsar 201 PC-controlled 19-channel electroencephalographic device was used to record EEG data. The operation took only two and a half hours for the placebo group. The VCPT is an expansion of the visual GO/NOGO paradigm with a third category and an additional two blocks. The objective was to immediately answer with a button press to each GO trial. For EEG analysis, independent component analysis (ICA) is a blind source separation method that use statistics of all orders to disentangle a chaotic mixture of distributions. It uses the waveforms of the source signal without knowing their timing to conduct blind separation. From a pool of 297 individual ERPs collected from healthy volunteers aged 18–45, seven separate

ERP components were extracted using linear algebra. The parts were chosen for their usefulness and strength, and their locations were determined with the help of sLORETA imaging and feature extraction. The key concept is that SVMs may be used to find the optimal separating hyperplane in a higher-dimensional space, while feature templates can be used to automatically produce a set of features. Group differences in psychometric test scores, behavioural outcomes, and independent component amplitudes were evaluated using the t-test, and cross-validation helped determine which features and classifiers to utilise. The middle parietal cortex, the supplementary motor cortex, and the anterior cingulate cortex contain the three set-specific components, while the lingual gyrus in the occipital lobe houses the three set-independent components. The current study used a modified version of the visual two-stimulus GO/NOGO task to investigate whether or not features of independent ERP components can be used to reliably differentiate between patients with ADHD and control subjects. To best classify data, an automated system considered a total of twenty criteria, but ultimately settled on five: extremely high latency values within small time frames. These characteristics may be linked to the disabling of a preprogrammed reaction triggered by the observation of an aberrant occurrence. The ADHD ERP literature infrequently reports latency variations in the BA 6 novelty component, as well as the BA 18, BA 25, and BA 5 components. [5]

Chapter 3

Dataset, Data Analysis & Data Pre-processing

3.1 Dataset Description

The dataset that we work on is one of the most important components in our thesis, there are two set of data. So, we collected it from [18] which is a T1 weighted data for brain researchers who want to do analytical studies, and the second one is from [28]. The collection is mostly about analytical data that help researchers learn more about how the brain works and how it connects to the other parts of the brain. In the first dataset, it comprises analytical data from 35 subjects with ADHD, and 44 subjects without ADHD who all performed selective attention and response inhibition tasks Also, this analytical has been used in many other researchs [30]. We will mostly focusing on our first dataset as it has higher resolution and more clear slices.

Label	**Participant**
ADHD	44
Non ADHD	35

Table 3.1: Number of participant having ADHD and Non ADHD for 79-participants

Label	**Participant**
ADHD	12
Non ADHD	26

Table 3.2: Number of participant having ADHD and Non ADHD for 38-participants

From Table 3.1 & Table 3.2 we have two data set. the Table 3.1 has a total of 79 participant in which 44 are diagnosed as ADHD and other 35 subjects are normal people who do not have any ADHD symptoms. In the Table 3.2, it is shown that out of 38 participants only 12 of them has ADHD symptoms and rest of the 26 do not have ADHD. For the second dataset it seems the data is very imbalanced, it do not have a proper ration of ADHD and Non ADHD subjects.

The dataset contains sMRI data that has been put through a number of analytical steps to improve its quality and get rid of any errors or noise. These preparation steps might include correcting motion, normalizing space, smoothing, and getting rid of sources of bodily noise. By using these analytical methods on the information,it has given us a high-resolution structural image of the brain that is commonly used for anatomical analysis.

The dataset's availability of statistical maps is one of its main analytical characteristics. Based on the sMRI data, these images show patterns of brain activity or connections. On the preprocessed sMRI data, statistical analyses like general linear models or correlation analyses are often used to find brain areas or networks that respond significantly to certain testing tasks or inputs. The mathematical pictures that came out of it let researchers see and study these activity patterns across the brain.

Also, it is a time series data for an area of interest (ROI). ROIs are parts of the brain or other parts of the body that are important to a study. The sMRI data from these ROIs is taken and processed to get time series data that shows how brain activity changes over time in these areas.The collection could also include measures or metrics that are based on the data and show how certain parts of the brain work or link. These derived measures can include things like functional connectivity, which looks at how activities in different parts of the brain are related or work together, or network properties, which describe how brain networks are set up and how well they work. These analytical measures give important information about how the brain works and can be used to look for links between brain activity and cognitive skills.

In short, the dataset has a lot of analytical parts that helped us to learn more about how the brain works and connects. Statistical maps, ROI time series data, and derived measures give us information about structural patterns, functional connections, and network features.

3.2 Data Analysis

Exploratory Data analysis or EDA is a fundamental and crucial part for analyzing data. As our thesis is mostly focused on this data, we need to evaluate and preprocess it thoroughly so that we can prepare it for model fit.

The dataset that we have is raw structural MRI data of each subject, that is t1 weighted. It means that the data is about the structure of the brain. To further analyze the data we opened & observed each MRI file to understand what kind of parameter and dimension it has. After analysing the files with the help of Nibable library, which is used to open Nifti file format, it loads and visualizes a 3D MRI volume, and the header information provides essential details about the scan acquisition and voxel dimensions, which are crucial for proper spatial analysis and interpretation of the MRI data. The header contains metadata about the MRI acquisition, such as pixel dimensions, slice thickness, scan duration, and spatial transformations.

sizeof_hdr
dim
pixdim
slice_duration
sform_code
qform_code

Table 3.3: Metadata found after scanning header information

But for our work we need to focus on some certain metadata which might help us in our work and those cab be found in Table 3.3
Inside our MRI data we have a 3 dimensional brain image, which means there are three axes. In brain MRI, the axis refers to the orientation of the imaging slices with respect to the three-dimensional coordinate system of the brain. There are typically three main imaging planes used in brain MRI, Axial Plane, Sagittal Plane and Coronal Plane.

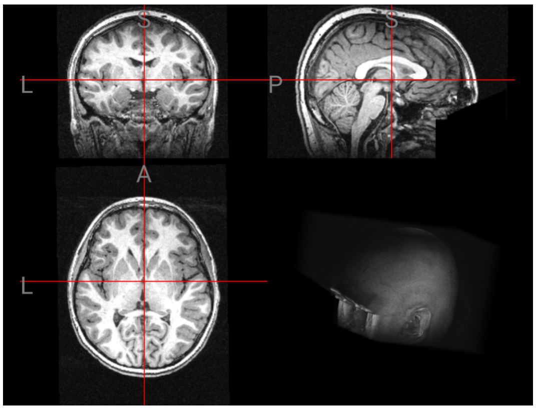

Figure 3.1: Raw MRI slices for data analysis

By acquiring imaging slices in these three planes, we can visualize and analyze different brain structures and pathologies from various perspectives. The axial plane is the most commonly used imaging plane in routine brain MRI, as it provides a comprehensive assessment of the brain's anatomy and is ideal for detecting a wide range of neurological conditions. However, the use of other planes such as sagittal and coronal planes is essential in detection of ADHD for our work.

3.3 Data Preprocessing

The MRI data that we have worked on was initially raw data in NifTi file format. Inside the Nifti file there were three dimensional brain images and each of the axes had corresponding slices of the subject's brain. The shape of each slice was (256, 256, 160). For our convenience first we processed the raw data to convert each Nifti file to image slices according to their dimension. For the x and y axis the total number of image slices are 256 and for the z axis there are 160 slices. For our work we only need 62 middle slices. As, while loading and saving those image slices we did drop the rest of the slices since they were blank or the image quality was not clear. for each image we have total 186 slices of middle images.

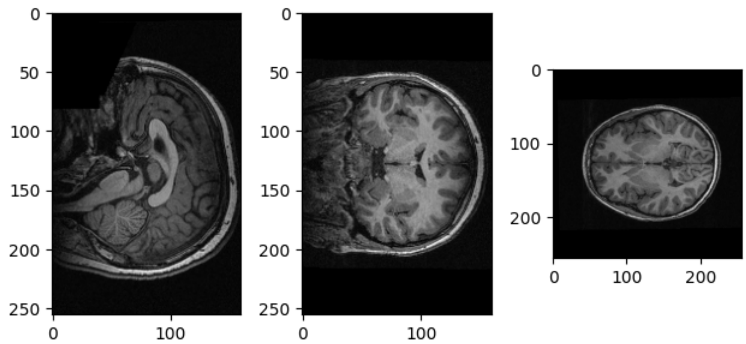

Figure 3.2: Scanned array's middle slices with proper aspect ratio

The middle part of the image slices are the most clear and brightest, as we are working with structural MRI data we need every detail of the image so that our CNN models can extract the features and train it properly based on those features. After successfully converting the files to images we applied some image processing to normalize & resize the image to 64 x 64 from 255 x 255.

After done with the image processing now we split our data to fit in our model. Table 3.4 shows how the data has been splitted.

Splitted Data	Image Data
training	7737
validation	2418
testing	1935

Table 3.4: Data splitted for training, testing and validating

here, first of all we are splitting the data into two part one is for training 80% and one is for testing which is 20%. the testing data will be used later for pedicting the model. then, we again splitted the training data into two parts, where we had again 80% traing data and for validation data this time 20%. here, the validation data will be evaluating the model while training the model with train dataset.

Chapter 4

Methodology, Architecture & Model Specification

4.1 Overview of the Proposed System

In order to begin using this system, the sMRI dataset was first placed into our device. After that, we took the structural MRI data and turned it into image data, after which we put specific image datas into our local storage. Following that, we imported the picture data from local storage and did some simple image processing like, reshaping the image size and normalizing the data so that it can fit in our model. After that we splitted the data for use in training and testing respectively. The data that we had previously divided for training was then split for validation and training purposes. After that, we carried out the model implementation by making use of the validation and training data in a partitioned format. During the implementation, we made use of three pre-trained models in addition to two models that were not pre-trained. The pre-trained models that we used are as follows: Densely-connected Convolutional Network (DenseNet-121), Visual Geometry Group (VGG-16), and Residual Neural Network (ResNet-50). Convolutional neural network (CNN) and convolutional long short-term memory (Conv LSTM) are examples of the non-pretrained models that we utilized. After that, we utilized a method called ensemble learning to construct a model called an ensemble model by analyzing the results of the best three models which has the best accuracy among the five models that we had previously used. A forecast was assessed based on the results of a majority vote, and it provided the intended outcome. A detailed overview has been shown in the the diagram below.

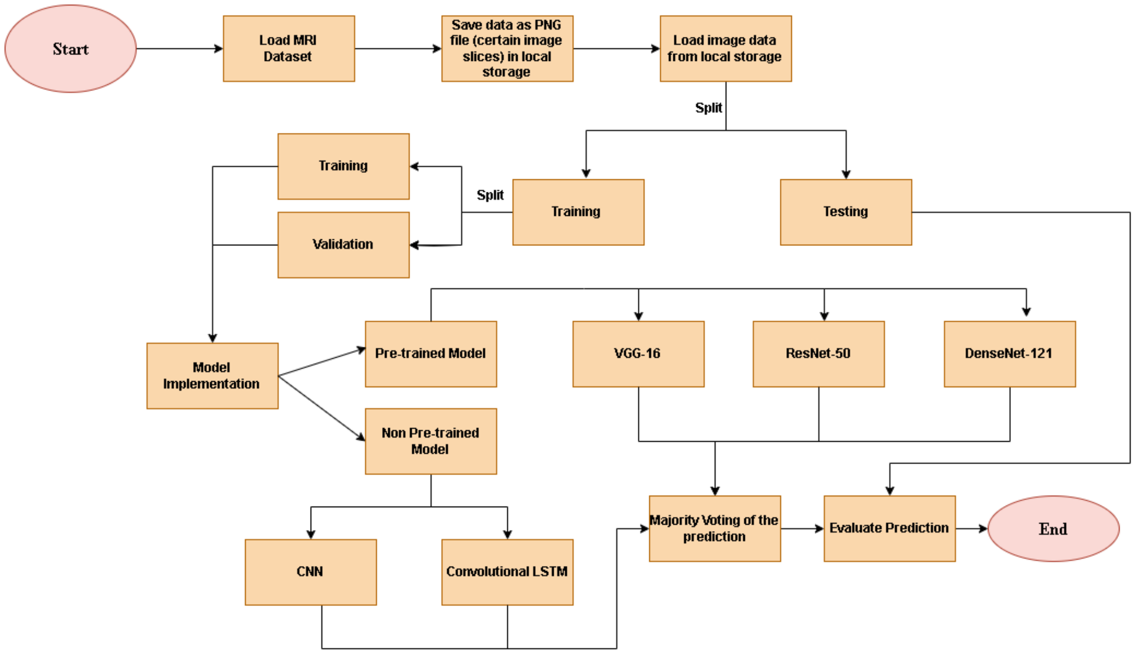

Figure 4.1: Flowchart of the proposed System

4.2 Model Description

In this study project, a wide range of deep learning architectures, such as Convolutional Neural Networks (CNN), ConvLstm, VGG16, Resnet50, and DenseNet-121, were carefully used. This smart choice of neural network models was based on a strategic goal: to face and overcome the many challenges that are inherent to the study domain. As CNN is the best model for image data feature extraction, we used CNN for getting efficient outcome from the image data feature extraction. So for The deployment of this diverse collection of architectures was not arbitrary; rather, it was a calculated decision made to leverage their individual strengths and fine-tune them to address particular facets of the complex research problem at hand. The study, which was conducted with rigor and accuracy, aimed not only to demonstrate the efficacy of these sophisticated deep learning approaches, but also to provide a nuanced understanding of their respective contributions, thereby enhancing the repertoire of machine learning techniques applicable to a wide variety of domains and scenarios.

4.2.1 VGG-16

VGG16 (Visual Geometry Group 16) is a deep convolutional neural network (CNN) architecture that was introduced by the Visual Geometry Group at the University of Oxford in 2014. It is one of the pioneering CNN models and has had a significant impact on the field of computer vision and deep learning.VGG16 is a deep convolutional neural network (CNN) that consists of multiple layers, including convolutional layers, max-pooling layers, and fully connected layers. It works in MRI data by processing the input MRI image through these layers to extract relevant features and make predictions.[23] Using VGG16 in MRI improves MRI reconstruction due to its feature extraction ability, transfer learning advantages, capturing high-level representations from pre-trained weights, and robustness in handling complex image data

(GALAL & YAHYA, 2021).[29][15]

VGG-16 architecture

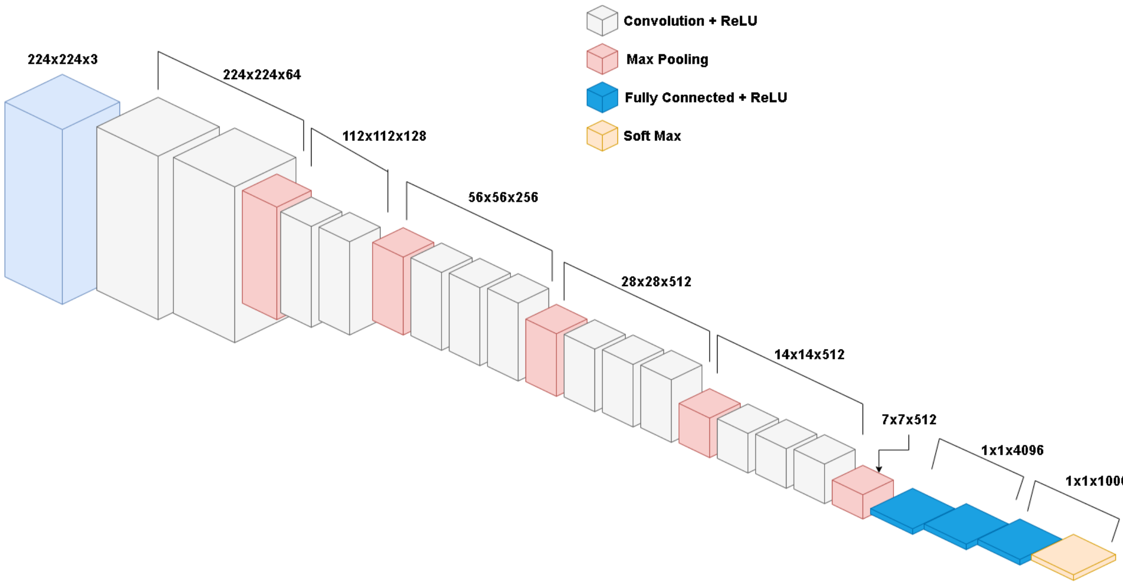

Figure 4.2: Architecture Of VGG-16 Model

Input Layer

Input images are typically of size 224x224 pixels (RGB).

Convolutional Blocks

The network is made up of 13 convolutional layers that are paired into 5 convolutional blocks. Each block comprises multiple 3x3 convolutional layers followed by a max-pooling layer. Increasing the number of filters in each convolutional layer promotes the network's ability to learn advanced characteristics. After each layer of convolution processing, the Rectified Linear Unit, also known as ReLU, is employed as the activation function.[29][15]

Fully Connected Layers

After the curved blocks, there are three additional layers that are all linked.[23] Each of the first two levels, which are fully related, has 4096 units. The output layer, which is the final fully linked layer, has as many units to be there are classes within the classification job. To get class odds, the function of softmax activation is used on the output stage.[29][15]

Dropout Layers

To reduce overfitting, dropout layers with a dropout rate of 0.5 are added after the first two fully connected layers.
Batch normalization is used to get the stimulation in each layer to appear the identical, which helps train the network to enhance its ability to learn novel concepts.

The cross-entropy loss function, that is often used in classification tasks, is utilized to train the network. [23]It may be trained on large image datasets such as ImageNet and then fine-tuned for specific tasks. VGG16 has been used extensively as an image classification algorithm, object the detector, and image segmenter. It has additionally been employed as a baseline design for many computer vision tasks. But it's crucial to keep in mind that VGG16 is fairly complex and has a lot of parameters. This renders it harder to compute while using up more ram than more modern designs such as ResNet and EfficientNet.[29]

VGG16, or Visual Geometry Group 16, is a convolutional neural network architecture used for image classification. There are 16 layers total, 13 of which are convolutional and 3 of which are fully linked. VGG16 uses tiny 3x3 convolutional filters all throughout the network to capture fine details in images. To reduce spatial dimensions and improve computing performance, it employs max-pooling layers. The subsequent layers are fully coupled, leading to a softmax classification layer. VGG16 is noted for its simplicity and deep architecture, which makes it useful for a variety of computer vision problems, while it has been outperformed in terms of performance by newer architectures such as ResNet and Inception.[15]

4.2.2 ResNet-50

ResNet-50, short for ”Residual Network-50,” is a deep convolutional neural network (CNN) architecture that was introduced by Kaiming He et al. in their 2015 paper titled ”Deep Residual Learning for Image Recognition” [4] .[17] It is part of the ResNet family, a series of neural networks that pioneered the use of residual connections or skip connections to address the vanishing gradient problem in very deep networks.Also , It’s a pre-trained model.[24]

ResNet-50 architecture

Input

ResNet-50 takes an RGB image as input, typically with dimensions 224x224 pixels.

Convolutional Layers

The network begins with a set of convolutional layers that pull out features. These layers use small convolutional filters to capture different parts of the input picture.[17]

Residual Blocks

The main new thing about ResNet-50 is the residue block. Instead of just stacking layers in order, ResNet has skip connections or shortcuts that go around one or more levels. This lets the network learn residual functions, which makes it easy to train very deep networks. Usually, each residual block is made up of a few convolutional layers and then a few shortcut links.

Pooling Layers

After removing some leftover blocks, max-pooling layers are used to make the feature maps smaller while keeping the important features.[17]

Fully Connected Layers

At the end of the network, fully joined layers are used for the final classification. These layers take the high-level traits found by the previous layers and map them to the output classes. For ResNet-50, this often leads to 1,000 output units for putting images into 1,000 different groups, like in the ImageNet Large Scale Visual Recognition Challenge.[24]

Softmax Activation

The last layer uses the softmax activation function to give each class a chance score. This shows how the odds are spread out over all the different classes. Resnet first introduced the concept of skip connection. Skip connection means Add the original input to the output of the convolution block.[17] That means Adding the original

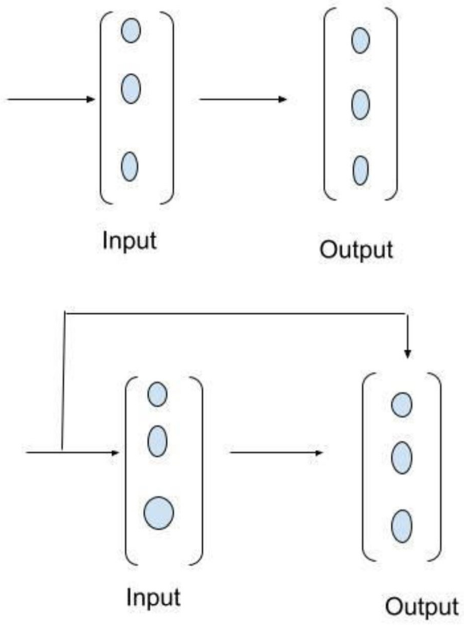

Figure 4.3: Architecture Of ResNet-50 Model

input to the output of a convolutional graph.Through the Skip connection Resnet solves the vanishing gradient because some layers are being skipped so we will not reach the small value.[24]

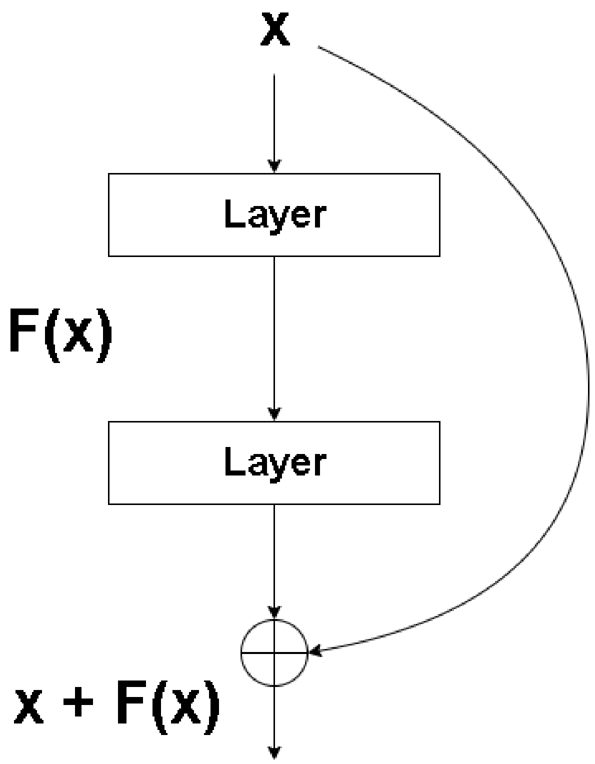

Figure 4.4: Architecture Of ResNet-50 Model

Here x = Input and F(x) = loss function.
Y = F(x) + x (because we want input and output should be equal).
make F(x) = 0 so that we can make Y = X.

The main idea is add the actual input into your loss function and try to make F(x) = 0 so that we can make Y = X (Y = X because we want to increase the overall accuracy).

ResNet-50 is a deep convolutional neural network design that is used for image classification. It works by putting layers of convolutional processes on top of each other to find hierarchical features in images. The biggest change is the addition of residual blocks with skip links that let the network learn residual functions and make it easier to train very deep networks.[24] These skip connections allow for the direct passage of information throughout the network and alleviate the vanishing gradient problem, resulting in enhanced training and greater precision. Then, ResNet-50 uses pooling layers to lower the number of spatial dimensions. Next, fully connected layers are used to classify the images using a softmax activation function. Because of how well this design trains deep networks, it has become a key part of computer vision applications.[17]

4.2.3 DenseNet-121

DenseNet-121 is a popular convolutional neural network (CNN) architecture used for computer vision tasks, such as image classification. We use it to handle the problem of vanishing gradient aslo it provide high Accuracy. It was introduced by Huang et al. in the paper "Densely Connected Convolutional Networks" in 2017.[22]

DenseNet-121 Architecture

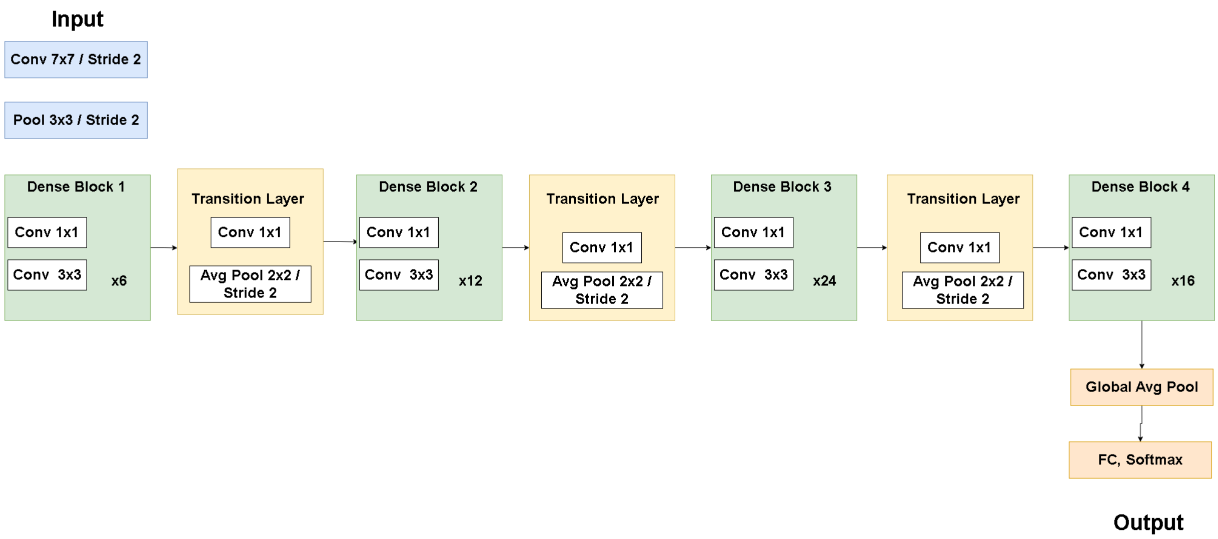

Figure 4.5: Architecture Of DenseNet-121 Model

Input Layer

The network takes an input image typically of size 224x224 pixels with three color channels (RGB).

DenseNet-121 Architecture

Initial Convolutional Layer

DenseNet-121 starts with an initial convolutional layer with 64 filters (kernels) of size 7x7, followed by batch normalization and a ReLU activation function. This layer is used to capture low-level features from the input image.[20]

Dense Blocks

DenseNet-121 consists of four dense blocks. Each dense block contains several convolutional layers (usually 6) that have 1x1 and 3x3 convolutional filters. Batch normalization and ReLU activation functions are applied after each convolutional layer within a dense block.[22]

Transition Layers

Between the dense blocks, there are transition layers that consist of a 1x1 convolutional layer followed by 2x2 average pooling. The purpose of these transition layers is to reduce the spatial dimensions (width and height) of the feature maps while increasing the number of filters gradually.[22]

Global Average Pooling

After the final dense block, a global average pooling layer is applied. This layer computes the average value for each feature map across its spatial dimensions, resulting in a fixed-size feature vector.[22][20]

Fully Connected Layer (Dense Layer)

The global average pooled features are connected to a dense layer with a softmax activation function at the output. This final layer produces class probabilities for image classification.DenseNet-121 specifically refers to a variant of DenseNet with 121 layers. Variants with different depths, such as DenseNet-169 and DenseNet-201, have also been developed. These models are often pre-trained on large image datasets like ImageNet and fine-tuned for specific image classification tasks.[22]

DenseNet-121 is a deep convolutional neural network design that is used for image classification. It works because each layer is tightly linked to every other layer in a feedforward way. Unlike traditional convolutional networks, in which information flows sequentially from layer to layer, DenseNet encourages direct connections between all layers. This lets features be reused and gradients run throughout the network. Each layer gets the feature maps from all the layers that came before it and sends its own feature maps to all the layers that come after it. This dense connection architecture promotes feature propagation, alleviates the vanishing gradient problem, and yields a very efficient and accurate model for numerous image recognition applications.[20]

4.2.4 CNN

CNN stands for Convolutional Neural Network. It is a type of deep learning algorithm used primarily for image and video recognition tasks, although it has also been applied to various other tasks involving sequential data.[14]

CNN architecture

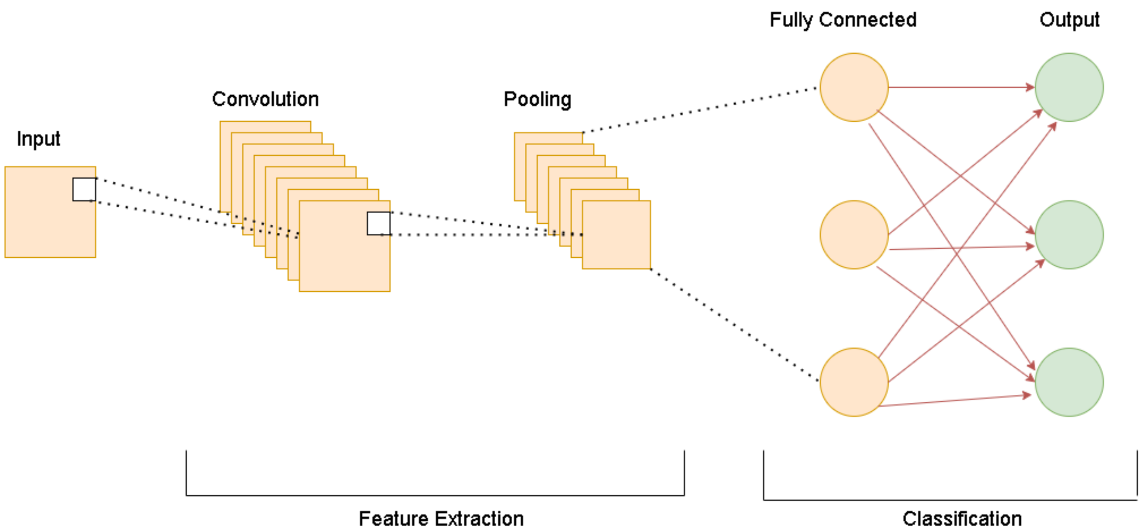

Figure 4.6: Architecture Of CNN Model

During the course of many years, advances in both speed and precision were made to the structure of deep convolution neural networks (CNNs). Novel structures, such as VGG, GoogLeNet, ResNet, DenseNet, ResNeXt, SE-Net, and autonomous neutral architectural search, have each considerably improved the network's accuracy in classification since AlexNet's seminal work . [14] The precision and efficacy of image categorization has been greatly advanced by the aforementioned architecture. [13] A summary of these seminal CNN architectures is given below:

Input Layer

CNNs begin by creating an input layer that takes an image as its input. The image can be seen as a grid for pixel values, with every pixel's power displaying the color or grayscale information.[14][13]

Convolutional Layers

These are the basic components of CNNs. The input image is handled by filters, which are commonly referred to as kernels, in convolutional layers. These filters move throughout the picture to do convolutions in general which are used for picking out details such as edges, corners, and textures. Since each convolutional layer has many filters, it provides the network the capacity to learn multiple features at the same time.[14][13]

Activation Function

After each convolution, an activation function (usually ReLU, which means Rectified Linear Unit) is attached to each element. It makes the network shorter, which helps it model intricate relationships in the data.[14][13]

Pooling Layers

After each convolution, an activation function (usually ReLU, which means Rectified Linear Unit) is attached to each element. It makes the network shorter, which helps it model intricate relationships in the data.Pooling layers follow convolutional layers. They downsample the feature maps gained from the previous layers, decreasing spatial dimensions. Max pooling as well as average pooling are popular methods to keep the most important information while decreasing computational complexity.[14][13]

Fully Connected Layers

These stages typically occur at the end of the network. They consist of neurons connecting to all neurons in the previous layer in order to produce a conventional neural network. These levels bring together high-level information to make predictions at the end.[14][13]

Output Layer

The output of the network is generated by the final layer that is fully linked. Depending on the task, the output layer might include one or multiple neurons that show classes in classification tasks or values in regression tasks. Activation functions are used, like softmax (for classification) or linear (for regression).[14]
In summary, CNNs excel in image-related tasks by automatically learning and extracting relevant features from the data through a combination of convolutional layers, activation functions, pooling layers, and fully connected layers. This hierarchical feature extraction, coupled with parameter learning, allows them to make accurate predictions in tasks like image classification, object detection, and more.[13]

Convolutional Neural Networks (CNNs) are a type of deep learning model that is made to handle and analyze structured grid-like data. The most common use of CNNs is to analyze images. Because of how they are built, CNNs are great at jobs like classifying images, finding objects in images, and separating images into parts. At the heart of a CNN are convolutional layers, which use a set of filters that can be taught to the data that comes in. These filters, also known as kernels, slide over the input image in a process called convolution, extracting features like edges, textures, and shapes at various spatial hierarchies. The next layers often have pooling layers, which decrease the size of the space and make the features even less clear. After a few of these convolutional and pooling layers, fully connected layers are used to describe high-level features and make decisions. CNNs are trained using labeled data and backpropagation. They have changed the way computer vision jobs are done by letting machines do things like image recognition and object tracking as well as humans. Their ability to automatically learn hierarchical features from data makes them powerful tools in many areas beyond computer vision, such as natural language processing and medical image analysis.[13][14]

Layer	Output Shape	Parameters
inputLayer	(None, 64, 64, 3)	0
Conv2D	(None, 64, 64, 128)	3568
Activation: ReLu	(None, 64, 64, 128)	0
MaxPooling2D	(None, 32, 32, 128)	0
Dropout: 30%	(None, 32, 32, 64)	0
Conv2D	(None, 32, 32, 64)	73792
Activation: ReLu	(None, 32, 32, 64)	0
MaxPooling2D	(None, 16, 16, 64)	0
Dropout: 30%	(None, 16, 16, 64)	0
Conv2D	(None, 16, 16, 32)	18464
Activation: ReLu	(None, 16, 16, 32)	0
MaxPooling2D	(None, 8, 8, 32)	0
Dropout: 30%	(None, 8, 8, 32)	0
Flatten	(None, 2048)	0
Dense: hiddenLayer1	(None, 64)	131136
Activation: ReLu	(None, 64)	0
Dropout: 50%	(None, 64)	0
Dense: hiddenLayer2	(None, 1)	65
Activation: Sigmoid	(None, 1)	0

Table 4.1: CNN Model Summary

4.2.5 Convolutional LSTM

In the paper "Convolutional LSTM: A deep learning approach for dynamic MRI reconstruction" by Yakkundi and Subha, Convolutional LSTM is described as a deep learning technique used for dynamic MRI (Magnetic Resonance Imaging) reconstruction.Convolutional LSTM combines the concepts of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks.[21][25] It utilizes convolutional operations to process spatial information and LSTM cells to handle sequential dependencies in the MRI data. The ConvLSTM (Convolutional Long Short-Term Memory) is a neural network architecture that combines the properties of convolutional neural networks (CNNs) and Long Short-Term Memory (LSTM) networks.[30] It is particularly useful for tasks that involve sequences of data with spatial dependencies, such as video processing, weather forecasting, and spatiotemporal data analysis. ConvLSTMs extend traditional LSTMs by incorporating convolutional operations to process input sequences in a more spatially aware manner.[26]

Convolutional LSTM architecture

Input

A ConvLSTM cell takes data in the shape of a sequence, which typically appears as a 3D tensor with dimensions (batch size, time steps, height, width, channels).

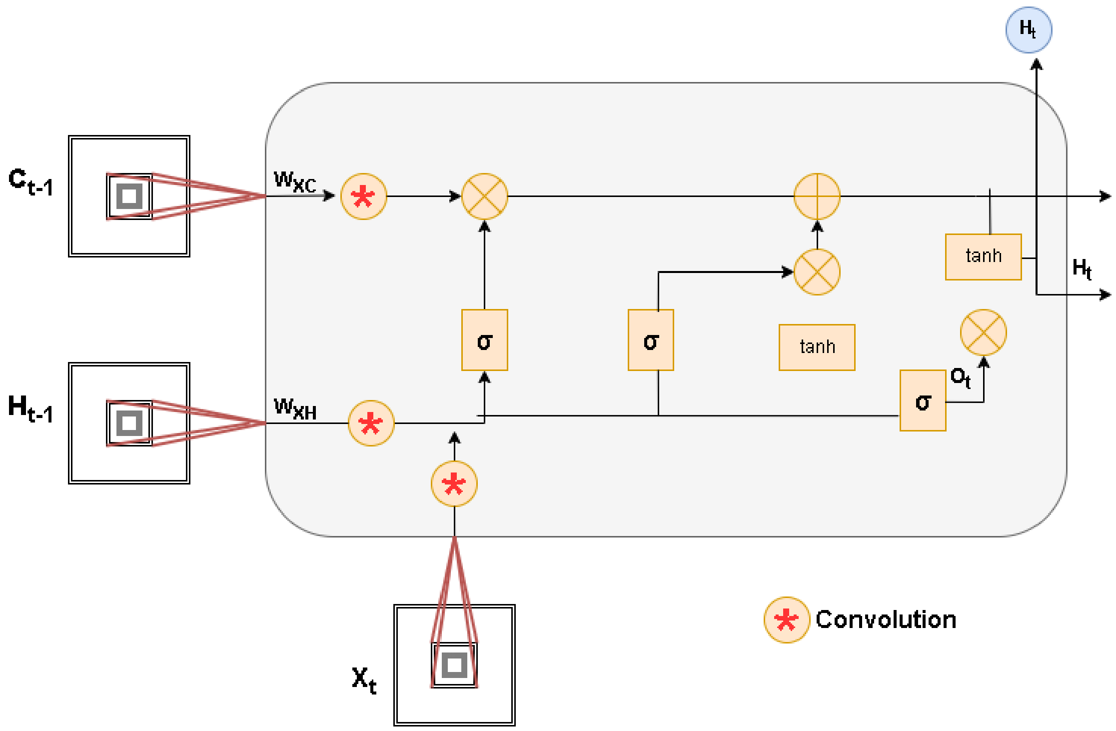

Figure 4.7: Architecture Of Convolutional LSTM Model

[21][25]Each time point in the sequence is represented by an image or snapshot, and every snapshot has a spatial structure which can be defined by its height, width, and canal dimensions.[26]

Convolutional Operations

ConvLSTM adds processes that do ”convolutions” to the LSTM cell. At each time step, those neural processes are done in the data that came in. The convolutional filtering methods assist in figuring out how the raw data is related to space.[26]

LSTM Cell State

ConvLSTM cells, like normal LSTM cells, have an internal state (often written ”C”) as well as a secret value (often written ”h”). These states change over time stages and are liable for keeping information and passing it on from one time step to another whilst taking into account either temporal as well as spatial relationships.[26]

Gate Mechanisms

Standard cells with LSTM have gates like the input gate, skip gate, and output gate. The ConvLSTM cells also have these gates. These gates control the flow of data into and out of the cell’s state. This lets the model choose which data to remember and what data to forget.[26]

Output

At each time step, an ConvLSTM cell gives a result that can be used to generate forecasts and passed to the next The ConvLSTM cell in a multi-layer architecture.[26]

The best thing regarding ConvLSTMs is that they may employ sequential information to find either temporal and spatial relationships. The convolutional neural networks let the model handle spatial features in each frame of the sequence, and the LSTM components assist in maintaining and spread information over time.ConvLSTMs are often used for tasks such as predicting what will happen in an audio recording, figuring out what is happening in a video, predicting when it will rain, as well as any other task where understanding spatial trends is important. Researchers have also made changes to ConvLSTMs, like Peephole ConvLSTMs and Stacked ConvLSTMs, in order to render them more effective doing specific tasks.[26]

ConvLSTM (Convolutional Long Short-Term Memory) is a neural network architecture that combines the spatial processing abilities of convolutional neural networks (CNNs) with the sequential memory retention of Long Short-Term Memory (LSTM) networks. It works by replacing the usual matrix multiplication operations in LSTM cells with convolutional operations. This enables ConvLSTM to capture both spatial and temporal dependencies in sequential data, making it suitable for tasks such as video prediction, motion analysis, and spatiotemporal data modeling. In each ConvLSTM cell, input data is mixed with learned filters, and the resulting feature maps are used to change the cell's memory state and make the output. This mix of convolutional and recurrent operations makes it possible for ConvLSTM to handle and model spatiotemporal data in a wide range of applications in an efficient way.[21][25]

Layer	**Output Shape**	**Parameters**
inputLayer	(None, 64, 64, 3)	0
Conv2D	(None, 64, 64, 32)	896
Activation: ReLu	(None, 64, 64, 32)	0
MaxPooling2D	(None, 32, 32, 32)	0
Dropout: 30%	(None, 32, 32, 32)	0
Conv2D	(None, 32, 32, 64)	18496
Activation: ReLu	(None, 32, 32, 64)	0
MaxPooling2D	(None, 16, 16, 64)	0
Dropout: 30%	(None, 16, 16, 64)	0
Conv2D	(None, 16, 16, 128)	73856
Activation: ReLu	(None, 16, 16, 128)	0
MaxPooling2D	(None, 8, 8, 128)	0
Dropout: 30%	(None, 8, 8, 128)	0
TimeDistributed: Flatten	(None, 8, 1024)	0
LSTM	(None, 64)	278784
Dense: hiddenLayer1	(None, 64)	4160
Activation: ReLu	(None, 64)	0
Dropout: 50%	(None, 64)	0
Dense: hiddenLayer2	(None, 1)	65
Activation: Sigmoid	(None, 1)	0

Table 4.2: Conv LSTM Model Summary

4.3 Ensemble Learning Technique

Ensemble learning is a machine learning technique that combines the predictions of various distinct models to get a single, more precise prediction. The theory behind ensemble learning is that it can frequently get greater predicted performance by combining the prediction of multiple models that it can achieve better predictive performance than with any individual model.[1] For both classification and regression tasks, it is a potent method that is frequently employed in machine learning.[1][4]

Although there are other kinds of ensemble learning techniques, we used majority voting for the task at hand. This enables the ensemble learning to use information from other models' predictions and perform a majority voting, which helps it outperform other models in terms of prediction accuracy.[4]

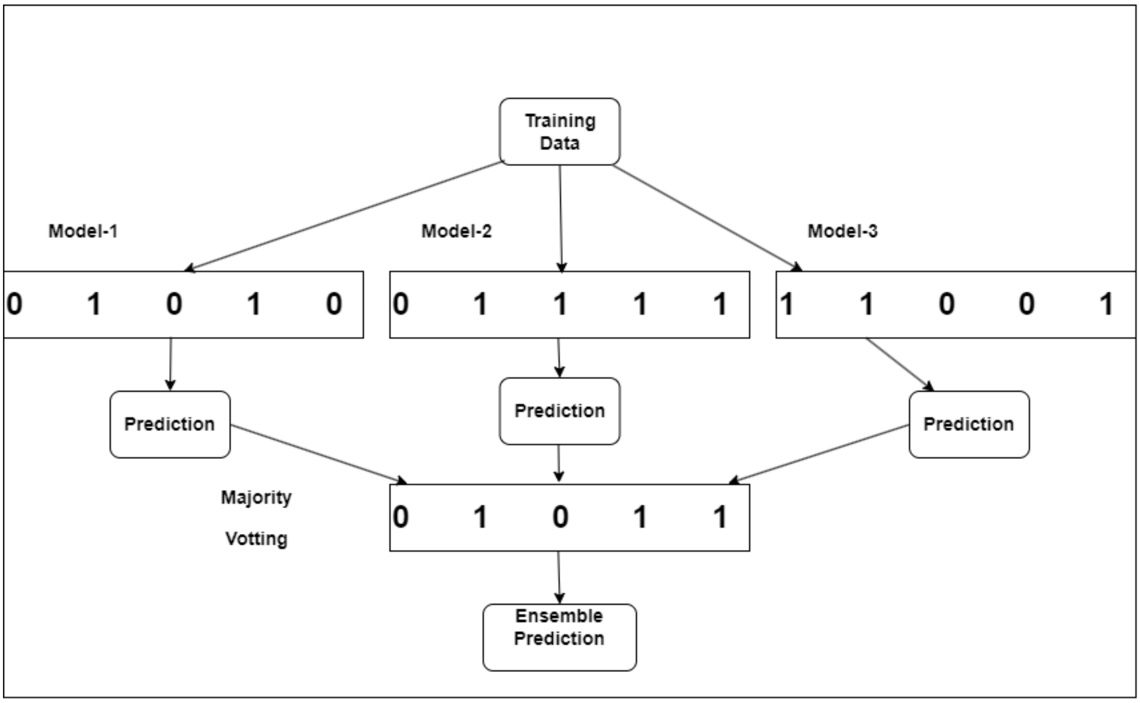

Figure 4.8: Ensemble Learning

Chapter 5

Result Analysis

5.1 Performance Evaluation Metrics

The Ensemble Learning technique that we used to detect ADHD from a MRI image has been evaluated through many performance evaluation metrics while doing our experiment. It has proven that the Ensemble Learning is performing well for our sMRI dataset and it has been evaluated by accuracy, precision, recall, F1-score, and confusion matrix. furthermore, the results demonstrate how effective the method is for classifying ADHD.

Accuracy: The accuracy metric represents a percentage of all samples that accurately identified the subject for having ADHD. The proposed method's accuracy has been used for evaluating our results using Equation 5.1

$$Acc = \frac{S_c}{T_c} \tag{5.1}$$

Here, Acc is the accuracy, Sc is for, correctly detected ADHD samples and lastly, Tc represents, total number of samples.

Precision: Precision is used to measure the accuracy of the positive predictions, which are made by classification models. This Precision score is used to evaluate the proposed method by using Equation 5.2

$$P_r = \frac{T_p}{T_p + F_p} \tag{5.2}$$

Here, Pr is the precision, Tp & Fp represents, true positive and true negative respectively.

Recall: Recall is a matrix that evaluates classification models where identifying positive instances is critical. Recall measures the models to correctly identify all ADHD positive subjects out of the total actual ADHD positive subjects. In the proposed method, the recall matrix has been calculated by using Equation 5.3

$$R_e = \frac{T_p}{T_p + F_n} \tag{5.3}$$

Here, Re is the Recall matrix, Tp & Fn are the true positive, and the false negative respectively.

F1-score: The F1-score metric is the harmonic mean of the precision metric and the recall metric. It displays a relative performance score for a model based on its recall and precision values. Based on the dataset, the F1-score was calculated using Equation 5.4

$$F_1 = 2 * \frac{P_r * R_e}{P_r + R_e} \tag{5.4}$$

Here, Pr & Re presents the precision and recall respectively.

Confusion Matrix: A confusion matrix is a visual output table, which is used to evaluate the performance of classification algorithms. It is a square matrix with the dimensions N*N, where N is the total number of classes to be classified. In a confusion matrix, columns represent actual classes and rows represent predicted classes. A confusion matrix demonstrates how the actual and anticipated ADHD subjects might be evaluated as a result.

ROC Curve: The ROC curve elaborates the performance of binary classification models, especially when there has to be a trade-off between correctly classifying positive instances and minimizing false alarms. It provides a visual representation of the model's ability to distinguish between the two classes under different classification thresholds. The TPR(True Positive Rate) is in the y-axis and the FPR(False positive rate) is in the x-axis.

5.2 Experimental Result Analysis

Model	Precision	Recall	F1-Score	Accuracy
VGG16	83%	60%	70%	80%
ResNet50	62%	60%	61%	64%
DenseNet121	69%	80%	74%	73%

Table 5.1: Evaluation Metrics Result for Propular Pre-trained CNN Models for 79 Participants

Model	Accuracy
VGG16	96%
ResNet50	59%
DenseNet121	90%

Table 5.2: Evaluation Metrics Result(accuracy only) for Propular Pre-trained CNN Models for 38 Participants

First of all, we have trained and tested our data with three pre-trained machine-learning models which are popular for medical analysis. As mentioned earlier they are, Visual Geometry Group (VGG16), Residual Neural Network (ResNet50) and Densely-connected Convolutional Network (DenseNet121). In Table 5.1 & Table 5.2 it shows the evaluation metrics result for 79 participants dataset and 38 participants dataset. From observing the results, apart from ResNet50 other models perform well for both dataset. Furthermore, the F1-Score of ResNet was very low compared to other models. F1-Score is an excellent technique for binary classification problems and evaluating machine learning model performance. As, it combines both precision and recall into a single value so that there should be no imbalance in the data. Beside that, DenseNet121 has higher F1-score than VGG16, but VGG16 wll be a better choice for ensemble learning, as it has higher accuracy score than DenseNet121. Second, the Table 5.2 only shows the accuracy result which was performed for 38 participant dataset. Here, from the accuracy it can be evaluated that ResNet50 is not suited for this dataset, where VGG16 and DenseNet121 is performing how a ML model should work. Even though in the 38 participant dataset we have better accuracy than the 79 participant dataset, 79 participants will be used for further experiment because the 38 participant has recurrent dataset in the testing and training set. So, that is why we are getting a higher rate of accuracy, as same data is passing on the test dataset and train dataset.

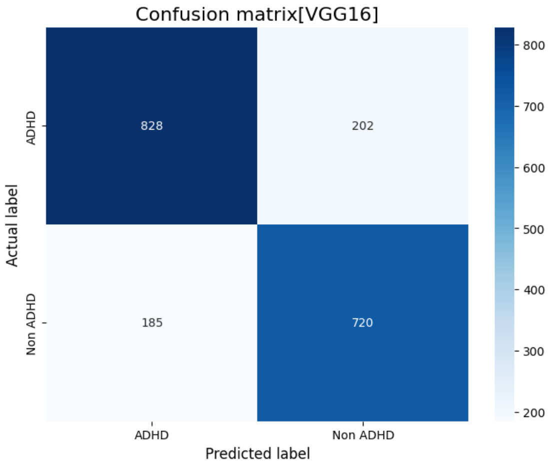

Figure 5.1: VGG16 Confusion Matrix

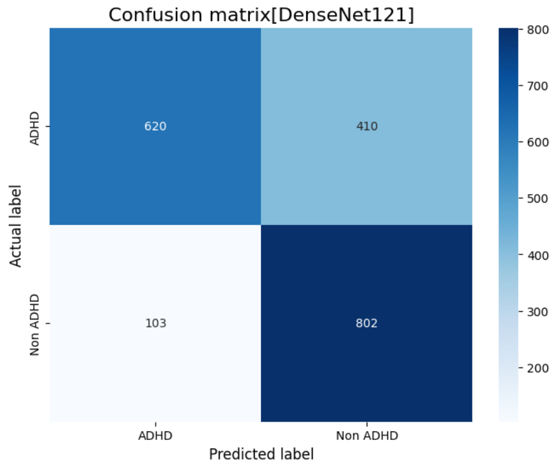

Figure 5.2: DenseNet121 Confusion Matrix

As mentioned earlier, the 38-participants will not be used so the main focus for further experiment will be the 79-participant dataset. The confusion matrix for Vgg16 and DenseNet121 in Figure 5.1 & Figure 5.2 respectively illustrates how many images they accurately predicted. Out of 1030 ADHD photos, DenseNet121 predicted 620 as having ADHD; the remaining 410 were predicted as not having ADHD but were really given the classification. This shows that the model correctly predicted 80% of the real ADHD cases. That is the recall score, but it also suggests that there were some false negatives - images which were actually ADHD but were not predicted as such. For Non ADHD, DenseNet121 predicted 802 images correctly but it identified 103 images ADHD out of 905 images. It means that out of all the images predicted as ADHD by DenseNet121, 69% were correct. While this is a good precision score, it also suggests that there were some false positives - images predicted as ADHD but were not. Since, the model is identifying a great number of images as false negative, its accuracy will be low because accuracy is the ratio of correctly predicted instances divided by total number of images. Here, images are the instances.

On the other hand, VGG16 is also a CNN based pre-trained model and its confusion matrix is shown in the figure-1, where out of 1030 ADHD images it predicted 828 to be ADHD and 202 images as Non ADHD, that means it has a high precision score of 83% and the model most of the time were correct. For recall Vgg16 has a 60% recall score, that means there are many false negatives - instances that were actually ADHD but were not predicted as such. Since the model has an accuracy of 80% it suggests that the VGG16 model is giving overall correct predictions on the data. Both VGG16 and DenseNet121 have performed excellently on our dataset but there are few differences like, VGG16 had higher precision than DenseNet121, so when it predicts ADHD it predicts correctly. Then again, recall has a higher score of 80% that makes the model to identify actual ADHD cases with fewer false negatives. The F1 score for both models are almost equal but DenseNet121 is slightly better than VGG16. But in terms of overall prediction for the dataset VGG16 performs best among all three models.

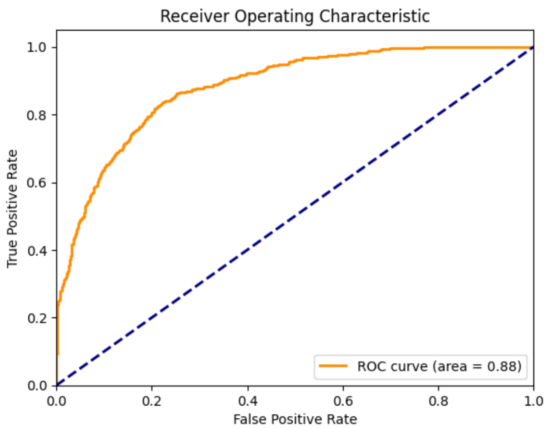

Figure 5.3: VGG16 Receiver Operating Characteristics (ROC Curve)

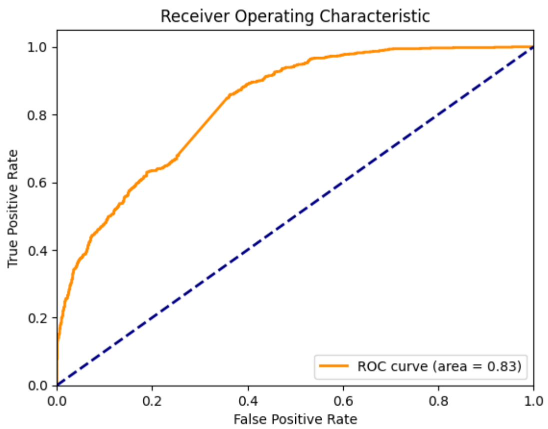

Figure 5.4: DenseNet121 Receiver Operating Characteristics (ROC Curve)

In Figure 5.4 and Figure 5.3 the Receiver Operating Characteristics (ROC) Curve indicates that there is a threshold starting from 0 to 1. The Y-axis is True Positive Rate (TPR) and X-axis is False Positive Rate (FPR). The ROC curve is an excellent tool for evaluating how the discrimination threshold of the model affects the trade-off between the true positive rate and the false positive rate. It describes the model's ability to distinguish between the two classes. A better discriminative model is typically indicated by a curve that is closer to the top-left corner. In this experiment, it shows that the VGG16 ROC curve is closer to the left corner than the ROC curve of DenseNet121. So, it can be confirmed that VGG16 has better performance to identify between two classes in a binary model. Moreover, VGG16 has been selected for ensemble techniques to perform better in ensemble learning methods. As for DenseNet121 it needs more improvement to identify between two classes.

Model	Precision	Recall	F1-Score	Accuracy	Loss
Conv_LSTM	93%	88%	91%	93%	0.20
CNN	92%	98%	95%	95%	0.14

Table 5.3: Evaluation Metrics Result for Convolutional LSTM and CNN

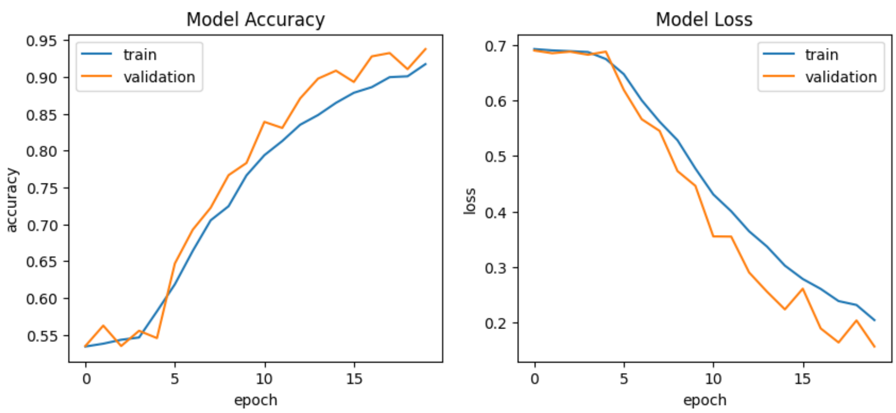

Figure 5.5: Accuracy & Loss curve for Conv LSTM

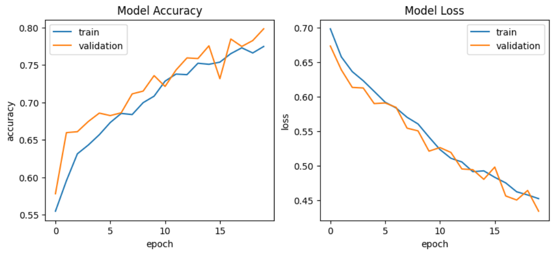

Figure 5.6: Accuracy & Loss curve for VGG16

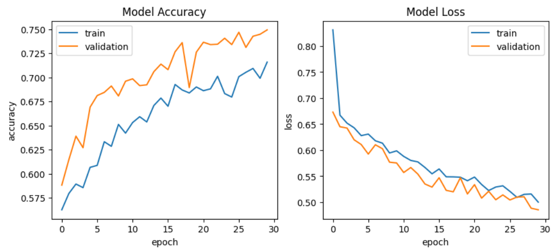

Figure 5.7: Accuracy & Loss curve for DenseNet121

In Table 5.3, it is shown that the evaluation metrics result for Convolutional LSTM and CNN. From there it can be confirmed that till now both of the above models performed better than any other pre-trained model so far. From Figure 5.5 it shows the train and validation accuracy & loss curve, which is definitely a better curve than the pre-trained model from Figure 5.6 Figure 5.7. All of these models run through 20 epochs. As declared earlier, if the curve of VGG16 is examined then it shows that after 5 epochs the validation curve almost touched the training curve and the rest of the time it spiked a lot and crossed with the training curve. Similar things did not occur in Figure-4, but the validation and training curve both spiked a lot which made the data training validating process unstable. In the case of Conv LSTM, its curve is more clear than the pre-trained models. For the first 5 epochs it seemed the model was struggling to train in the given data but after 5 epochs it got a continuous curve, after 10 epoch it had some spikes but it was not like the other two models. From Table 5.3, there is CNN that performed slightly better than Conv LSTM. In terms of accuracy both model achieved 90% accuracy which evaluates that the models are learning from the data.

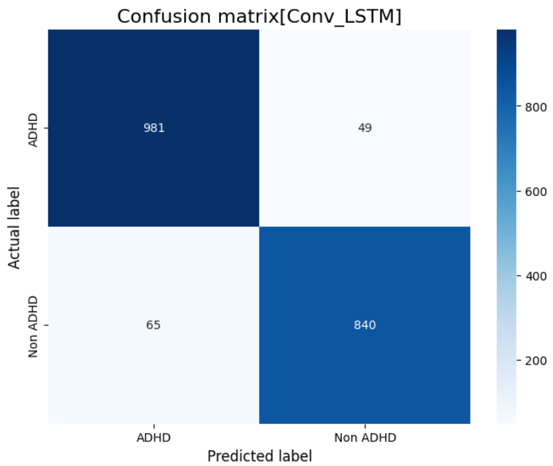

Figure 5.8: Conv_LSTM Confusion Matrix

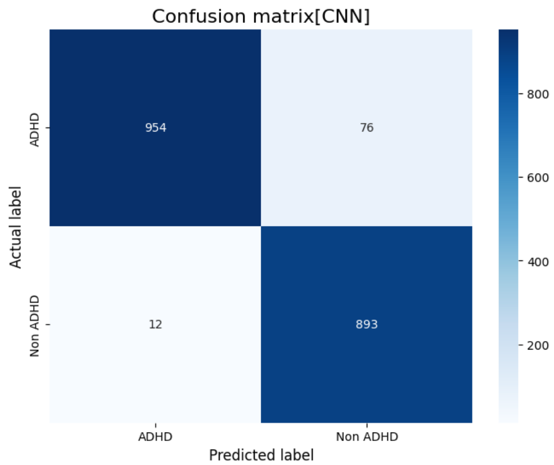

Figure 5.9: CNN Confusion Matrix

In order to evaluate more and understand about the models performance, some unseen data is provided as testing data that will assess the models true potential.

According to Figure 5.8, Figure 5.9, Figure 5.10, Figure 5.11, these Performance Evaluation Metrics provide some essential information about CNN and Conv LSTM. Even though there is slight difference between both models, there are other evaluation metrics that truly dictates which model has high performance. In Figure 5.8 and Figure 5.9, it provides confusion matrices of the models. First of all, Conv LSTM predicted 981 images as ADHD which were actually ADHD labelled as ADHD and only 22 images were predicted Non ADHD but it was actually ADHD. Then, for Non ADHD, the model predicted 840 images as Non ADHD but 65 of the images were identified as ADHD which were actually Non ADHD. From this information and the information given in Table 5.3, we can presume that Conv LSTM is avoiding false positives as the precision of the model is 93%. This means that out of all images 93% of the time it predicts actual ADHD. Then, the recall score is also very good for this model, as it can identify 88% of the images as ADHD from actual ADHD cases. The Conv LSTM model has an excellent degree of recall, showing that it captures a sizable fraction of positive cases, and a high level of precision, indicating that it minimises false positives. Its overall strong performance is further bolstered by the high accuracy and F1 score. Similarly, in Figure 5.9, the CNN model has close results like Conv LSTM, beside it performs a bit better. For CNN, it predicts 954 images as ADHD and 76 as Non ADHD but actually labelled as ADHD. Then, for Non ADHD it detects 893 images correctly and only 12 images as ADHD but were Non ADHD. Then, Precision for CNN is 92% where Conv LSTM has 93%. Just like that Recall is 98% F1-score is 95% and Accuracy is 95%. At first it may seem like the scores are superior to Conv LSTM. It is because both the models results are very close, So it can not be determined which model has the upper hand in terms of performance.

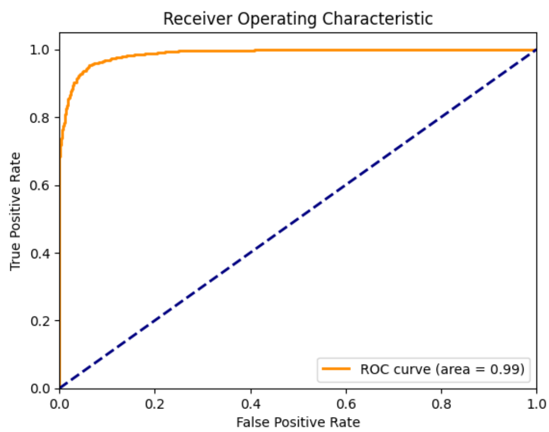

Figure 5.10: Conv_LSTM Receiver Operating Characteristics (ROC Curve)

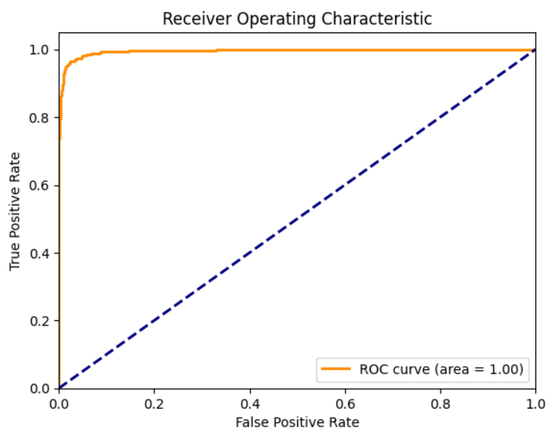

Figure 5.11: CNN Receiver Operating Characteristics (ROC Curve)

Even from Figure 5.10 & Figure 5.11, it shows the Receiver Operating Characteristics (ROC) curve, there is some marginally difference between those two models. As previously mentioned, in a ROC curve if a curve is more close to the left corner of the graph then it is most likely to have greater performance. For this case CNN is more close to the left corner as it can predict more true positives and less false negatives.

Model	Precision	Recall	F1-Score	Accuracy
Conv_LSTM	93%	88%	91%	93%
CNN	92%	98%	95%	95%
Ensemble Learning	96%	97%	97%	97%

Table 5.4: Evaluation Metrics Result for Convolutional LSTM, CNN & Ensemble Learning

However, after analysing all the models and comparing them with one another with the results, VGG16, Conv LSTM and CNN are the three best models out of five. From previous analysis it can not be determined which model is better CNN and Conv LSTM. So, we took another approach, where we will be using the best three models and do ensemble learning. In ensemble learning we will be doing majority voting from all three models output and try to get a better output. In Table 5.4 we implemented an Ensemble Learning model with the best three models and as it shows we got a higher accuracy with 97%, precision is 96%, recall and F1-score both are 97%.

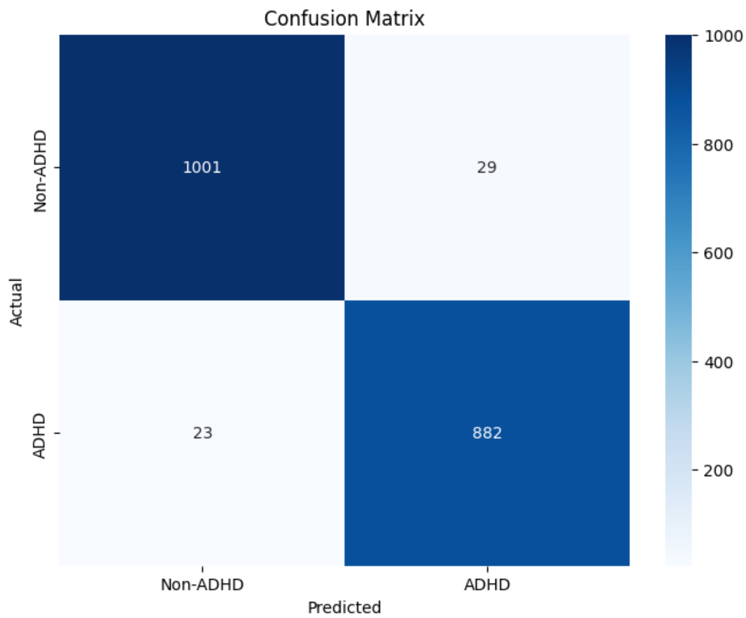

Figure 5.12: Ensemble Learning Confusion Matrix

From the confusion matrix of figure-1, the Ensemble Learning model has higher results than other three models. Only CNN has 98% recall score where Ensemble Learning has 97% recall. But otherwise, its accuracy and F1-score is better than others which indicates that Ensemble Learning has an excellent performance for our data.

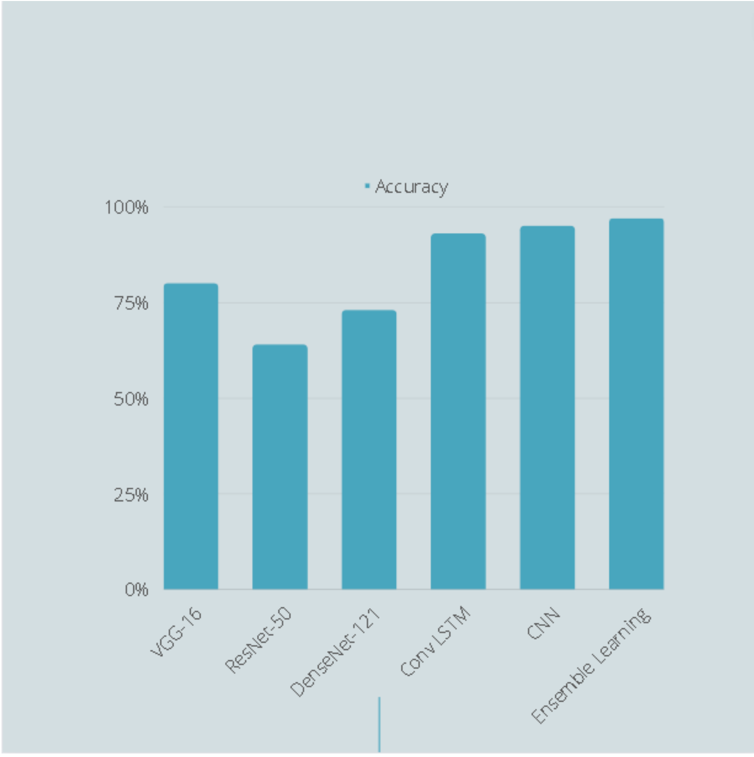

Figure 5.13: All five models accuracy

The above Figure 5.13 is showing a comparison of our model accuracy. From this figure it can be observed that our Ensemble Learning has the best output among other five model. As expected it should have the highest accuracy. If not then the Ensemble Learning was not implemented properly. However, it has been shown that with the help of majority voting we eshtublished a model that has 97% accuracy where CNN had 95.8% accuracy. So, our proposed model can helpful for Classifying ADHD detection.

5.3 Discussion

The experiment primarily centered on evaluating the performance of each model and thereafter doing a comparative analysis. In this thesis, we undertake a comprehensive examination and analysis of various models in order to obtain the optimal level of accuracy for our dataset. After picking the three finest models, we used the Ensemble Learning technique to them and successfully obtained the intended output. The findings of our study indicate that the implementation of Ensemble Learning resulted in a notable increase in accuracy. Prior to employing this technique, only the CNN and Conv LSTM models were able to achieve an accuracy rate over 90%. Despite achieving an accuracy rate of 80%, it was hypothesized that further improvements may be made by employing Ensemble Learning techniques to aggregate the prediction findings of VGG16. Following this, the classification of the classes is performed using the three models. Subsequently, the determination of which prediction to keep is made using a majority voting process, wherein the voting outcomes of the three models are taken into consideration. The class that receives the largest number of votes is selected, while other classes are discarded. The predicted classes for three models are [1, 1, 0, 1, 0], [1, 0, 1, 1, 1], and [1, 0, 0, 0, 1]. The following are the anticipated classifications for the VGG16, Conv LSTM, and CNN models, correspondingly. The prediction class for index 0 is determined to be 1, as all of the models in the analysis consistently return a value of 1 at their respective 0 index. However, in index 1, it appears that the first model has a value of 1, while the other two models have a value of 0 at index 1. As a result, the majority vote is in favor of 0 in this particular scenario. It can be estimated that, Ensemble Learning, when properly applied, is expected to consistently yield improved accuracy.

Chapter 6

Conclusion

In conclusion, by utilizing structural MRI data, our research has achieved important advancements in the field of ADHD (Attention Deficit Hyperactivity Disorder) diagnosis and research support. Our main goal was to create a reliable and accurate predictive model that can help physicians and psychiatrists with their research and contribute to the continuous search for fresh perspectives on ADHD. We have produced a number of noteworthy results through in-depth data analysis and the application of cutting-edge machine learning techniques. First off, based on structural MRI data, our model has shown to be remarkably accurate at predicting ADHD. It is highly effective at identifying people with ADHD while minimizing false positives and false negatives, according to the excellent precision and recall scores. Then, our methodology offers a useful resource for ADHD researchers in addition to helping with diagnosis. We can analyze MRI data more quickly and consistently by automating the classification process, which may one day lead to the discovery of novel structural patterns related to ADHD. In general, our findings signifies a significant advancement in the understanding and medicinal properties. We have developed a tool that not only helps with diagnosis but also advances knowledge of this difficult condition on a larger scientific scale by combining the strength of structural MRI data with machine learning.

6.1 Challenges

There is currently no such thing as an ADHD raw dataset accessible because there is no unique test or identification procedure that has been developed to diagnose ADHD as of yet. Because of this, it was necessary for us to get raw structural MRI datasets for our research. Because raw sMRI data are not immediately useable, we needed to transform the raw sMRI data into image data before we could use it in our work. After that, we used the image data in our work. The fact that all of the picture data for a particular patient consists of many slices was another difficulty that we had to work with. There is no correlation between having all the slices and having ADHD or not having ADHD. Therefore, it was necessary for us to choose particular slices for the picture data, and we made use of such slices in our research. In addition, the dataset was too large to load, which meant that we had to allot a significant amount of time to load the data and execute each epoc.

6.2 Future Work

During the course of our investigation, we made use of structural MRI data converting it into image slices. In the not too distant future, an effort will be made to utilize both functional and structural MRI data for the diagnosis of ADHD. Because it is dependent on the activity of the brain when executing a variety of tasks, including memory retrieval, it will produce better results. In future , it should be considered whether or not to explore how functional patterns in the brain evolve over time in people who have ADHD. When quick changes in brain function throughout a variety of tasks are able to be caught, it will be much simpler to determine which form of ADHD a patient suffers from. As a result of this identification, appropriate treatment and direction can be given.

Bibliography

[1] T. G. Dietterich, "Ensemble methods in machine learning," in *International workshop on multiple classifier systems*, Springer, 2000, pp. 1–15.

[2] J. Stevens, A. L. Quittner, J. B. Zuckerman, and S. Moore, "Behavioral inhibition, self-regulation of motivation, and working memory in children with attention deficit hyperactivity disorder," *Developmental neuropsychology*, vol. 21, no. 2, pp. 117–139, 2002.

[3] S. Durston, P. de Zeeuw, and W. G. Staal, "Imaging genetics in adhd: A focus on cognitive control," *Neuroscience & Biobehavioral Reviews*, vol. 33, no. 5, pp. 674–689, 2009.

[4] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction.* Springer, 2009, vol. 2.

[5] A. Mueller, G. Candrian, J. D. Kropotov, V. A. Ponomarev, and G.-M. Baschera, "Classification of adhd patients on the basis of independent erp components using a machine learning system," in *Nonlinear biomedical physics*, Springer, vol. 4, 2010, pp. 1–12.

[6] A. Nikolaidis and J. R. Gray, "Adhd and the drd4 exon iii 7-repeat polymorphism: An international meta-analysis," *Social cognitive and affective neuroscience*, vol. 5, no. 2-3, pp. 188–193, 2010.

[7] B. T. Felt, B. Biermann, J. G. Christner, P. Kochhar, and R. Van Harrison, "Diagnosis and management of adhd in children," *American Family Physician*, vol. 90, no. 7, pp. 456–464, 2014.

[8] R. Gwernan-Jones, D. A. Moore, R. Garside, *et al.*, "Adhd, parent perspectives and parent–teacher relationships: Grounds for conflict," *British Journal of Special Education*, vol. 42, no. 3, pp. 279–300, 2015.

[9] N. Wright, M. Moldavsky, J. Schneider, *et al.*, "Practitioner review: Pathways to care for adhd–a systematic review of barriers and facilitators," *Journal of Child Psychology and Psychiatry*, vol. 56, no. 6, pp. 598–617, 2015.

[10] M. Duda, R. Ma, N. Haber, and D. Wall, "Use of machine learning for behavioral distinction of autism and adhd," *Translational psychiatry*, vol. 6, no. 2, e732–e732, 2016.

[11] M. Sumathi and B. Poorna, "Prediction of mental health problems among children using machine learning techniques," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 1, 2016.

[12] S. Ziegler, M. L. Pedersen, A. M. Mowinckel, and G. Biele, "Modelling adhd: A review of adhd theories through their predictions for computational models of decision-making and reinforcement learning," *Neuroscience & Biobehavioral Reviews*, vol. 71, pp. 633–656, 2016.

[13] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo, "Multi-level wavelet-cnn for image restoration," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 773–782.

[14] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 116–131.

[15] Q. Guan, Y. Wang, B. Ping, *et al.*, "Deep convolutional neural network vgg-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: A pilot study," *Journal of Cancer*, vol. 10, no. 20, p. 4876, 2019.

[16] S. C. Guntuku, J. R. Ramsay, R. M. Merchant, and L. H. Ungar, "Language of adhd in adults on social media," *Journal of attention disorders*, vol. 23, no. 12, pp. 1475–1485, 2019.

[17] A. S. B. Reddy and D. S. Juliet, "Transfer learning with resnet-50 for malaria cell-image classification," in *2019 International Conference on Communication and Signal Processing (ICCSP)*, IEEE, 2019, pp. 0945–0949.

[18] J. R. Booth, G. Cooke, J. Gayda, *et al.*, *"working memory and reward in children with and without attention deficit hyperactivity disorder (adhd)"*, OpenNeuro, 2020. DOI: 10.18112/openneuro.ds002424.v1.1.1.

[19] H. Christiansen, M.-L. Chavanon, O. Hirsch, *et al.*, "Use of machine learning to classify adult adhd and other conditions based on the conners' adult adhd rating scales," *Scientific reports*, vol. 10, no. 1, pp. 1–10, 2020.

[20] D. Ezzat, H. A. Ella, *et al.*, "Gsa-densenet121-covid-19: A hybrid deep learning architecture for the diagnosis of covid-19 disease based on gravitational search optimization algorithm," *arXiv preprint arXiv:2004.05084*, 2020.

[21] M. Rahman, D. Islam, R. J. Mukti, and I. Saha, "A deep learning approach based on convolutional lstm for detecting diabetes," *Computational biology and chemistry*, vol. 88, p. 107 329, 2020.

[22] L. Sarker, M. M. Islam, T. Hannan, and Z. Ahmed, "Covid-densenet: A deep learning architecture to detect covid-19 from chest radiology images," *preprint*, vol. 2020050151, 2020.

[23] S. Srivastava, P. Kumar, V. Chaudhry, and A. Singh, "Detection of ovarian cyst in ultrasound images using fine-tuned vgg-16 deep learning network," *SN Computer Science*, vol. 1, pp. 1–8, 2020.

[24] D. Theckedath and R. Sedamkar, "Detecting affect states using vgg16, resnet50 and se-resnet50 networks," *SN Computer Science*, vol. 1, pp. 1–7, 2020.

[25] H. Wu, X. Ma, C.-H. H. Yang, and S. Liu, "Attention based bidirectional convolutional lstm for high-resolution radio tomographic imaging," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 4, pp. 1482–1486, 2020.

[26] S. V. Yakkundi and D. Subha, "Convolutional lstm: A deep learning approach for dynamic mri reconstruction," in *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, IEEE, 2020, pp. 1011–1015.

[27] A. Yasumura, M. Omori, A. Fukuda, *et al.*, "Applied machine learning method to predict children with adhd using prefrontal cortex activity: A multicenter study in japan," *Journal of Attention Disorders*, vol. 24, no. 14, pp. 2012–2020, 2020.

[28] J. R. Booth, D. D. Burman, D. R. Gitelman, *et al.*, *"response inhibition and selective attention in adults and children with and without adhd"*, OpenNeuro, 2021. DOI: 10.18112/openneuro.ds003500.v1.2.0.

[29] S. A. Y. AL-GALAL, I. F. T. ALSHAIKHLI, M. Abdulrazzaq, and R. HASSAN, "Brain tumor mri medical images classification with data augmentation by transfer learning of vgg16," *Journal of Engineering Science and Technology (JESTEC), Special Issue on ACSAT*, pp. 21–32, 2021.

[30] M. N. Lytle, D. D. Burman, and J. R. Booth, "A neuroimaging dataset on response inhibition and selective attention in adults and children with and without adhd," *Data in Brief*, vol. 37, p. 107 158, 2021.

[31] I. Öztekin, M. A. Finlayson, P. A. Graziano, and A. S. Dick, "Is there any incremental benefit to conducting neuroimaging and neurocognitive assessments in the diagnosis of adhd in young children? a machine learning investigation," *Developmental cognitive neuroscience*, vol. 49, p. 100 966, 2021.

[32] M. Maniruzzaman, J. Shin, and M. A. M. Hasan, "Predicting children with adhd using behavioral activity: A machine learning analysis," *Applied Sciences*, vol. 12, no. 5, p. 2737, 2022.

[33] P. Mikolas, A. Vahid, F. Bernardoni, *et al.*, "Training a machine learning classifier to identify adhd based on real-world clinical data from medical records," 2022.