Artificial Intelligence in ADHD Diagnosis Using CNN

Nada Ibrahim S. M. S. Kollah

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Computer Engineering

Eastern Mediterranean University
September 2022
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Ali Hakan Ulusoy
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science in Computer Engineering.

Prof. Dr. Hadi Işık Aybay
Chair, Department of Computer Engineering

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Science in Computer Engineering.

Assoc. Prof. Dr. Adnan Acan
Supervisor

Examining Committee

1. Assoc. Prof. Dr. Adnan Acan

2. Asst. Prof. Dr. Mehtap Köse Ulukök

3. Asst. Prof. Dr. Ahmet Ünveren

ABSTRACT

Modern medicine has a challenge in gathering, evaluating, and applying the vast amount of knowledge needed to address challenging clinical issues. The development of AI systems for medical applications has been linked to the development of medical AI. They are made to aid the physician in making a diagnosis, selecting a course of treatment, and anticipating results. Artificial Neural Networks (ANNs), fuzzy expert systems, evolutionary computation, and hybrid intelligent systems are some examples of such systems.

By processing EEG signals through two-dimensional colored picture transforms with GASF, the study aims to give early identification for Attention Deficit Hyperactivity Disorder (ADHD), one of the most prevalent neurobehavioral diseases. The EEG DATA FOR ADHD / CONTROL CHILDREN dataset of kids with ADHD disorder provided the EEG data.

ADHD is a hotly debated topic among medical professionals who contend that the condition is underdiagnosed and that many kids go undiagnosed. Therefore, it is crucial to bring such topic up and to find ways to help in easing the diagnosis procedure of this disease.

In this master's thesis, the 2D colored image data was trained and tested using the AlexNet deep learning model. Alexnet is an eight-layer Convolutional Neural Network (CNN) model. The main objective is to motivate researchers in medical image interpretation to extensively rely on CNNs in their analysis and diagnosis. In

the database, mentioned above, 16 channels were used. Among those 7 channels were used and those are Fp1, Fp2, F3, Fz, F4, P3, and P4.

Selected channels' GASF pictures were utilized to train the AlexNet model. For training and testing, the generated model achieved accuracy of 99% and 71.7%, respectively.

Keywords: Medical Condition, ADHD, Artificial Intelligence, Deep Learning, Neural Networks.

ÖZ

Modern tıp, zorlu klinik sorunları ele almak için gereken büyük miktarda bilgiyi toplamak, değerlendirmek ve uygulamakta zorlanmaktadır. Tıbbi uygulamalar için YZ sistemlerinin geliştirilmesi, tıbbi YZ'nin geliştirilmesiyle bağlantılıdır. Bu sistemler, hekimin teşhis koymasına, bir tedavi yöntemi seçmesine ve sonuçları tahmin etmesine yardımcı olmak için üretilmiştir. Yapay Sinir Ağları (YSA'lar), bulanık uzman sistemler, evrimsel hesaplama ve hibrit akıllı sistemler bu tür sistemlerin bazı örnekleridir.

Çalışma, EEG sinyallerini GASF ile iki boyutlu renkli resim dönüşümleri aracılığıyla işleyerek, en yaygın nöro-davranışsal hastalıklardan biri olan Dikkat Eksikliği Hiperaktivite Bozukluğu (DEHB) için erken teşhis sağlamayı amaçlamaktadır. DEHB / KONTROL ÇOCUKLARI İÇİN EEG VERİLERİ DEHB bozukluğu olan çocukların EEG verilerini sağlamıştır.

DEHB, bu hastalığın yeterince teşhis edilmediğini ve birçok çocuğa tanı konulmadığını iddia eden tıp uzmanları arasında tartışılan bir konudur. Bu nedenle, bu konuyu gündeme getirmek ve bu hastalığın teşhis prosedürünü kolaylaştırmaya yardımcı olacak yollar bulmak çok önemlidir.

Bu yüksek lisans tezinde, 2D renkli görüntü verileri AlexNet derin öğrenme modeli kullanılarak eğitilmiş ve test edilmiştir. Alexnet sekiz katmanlı bir Evrişimsel Sinir Ağı (CNN) modelidir. Temel amaç, tıbbi görüntü yorumlama alanındaki araştırmacıları, analiz ve teşhislerinde CNN'lere geniş ölçüde güvenmeye motive

etmektir. Yukarıda bahsedilen veritabanında 16 kanal kullanılmıştır. Bunların arasında 7 kanal kullanılmıştır ve bunlar Fp1, Fp2, F3, Fz, F4, P3 ve P4'tür.

Seçilen kanalların GASF resimleri AlexNet modelini eğitmek için kullanılmıştır. Eğitim ve test için, oluşturulan model sırasıyla %99 ve %70 doğruluk elde etmiştir.

Anahtar Kelimeler: Tıbbi Durum, DEHB, Yapay Zeka, Derin Öğrenme, Sinir Ağları.

DEDICATION

To my family

ACKNOWLEDGMENT

Special thanks go out to Assoc. Prof. Dr. Adnan Acan, my supervisor, for his visionary leadership and selfless efforts in ensuring the success of this research. It has been a fantastic opportunity to work with him, and his guidance has been a crucial component of my success in every phase of this project.

I would want to take this time to thank my parents for their prayers, support both emotionally and financially, advice, and care. I know I can't pay you back, but my goal is to let you know how much I value your assistance.

I want to thank everyone who has helped me along the path in whatever way. I would like to express my gratitude to all of my sisters and brother, friends, family, and coworkers, who are far too numerous to mention here. I'm speechless when it comes to thanking you for your invaluable assistance. I wish God's blessings on each and every one of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ADAM	Adaptive Moment Estimation
ADHD	Attention Deficit Hyperactivity Disorder
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
EEG	Electroencephalogram
GAF	Gramian Angular Field
GASF	Gramian Summation Angular Field
RGB	Red, green, and blue
RMSP	Root Mean Square Propagation

Chapter 1

INTRODUCTION

The most common neurobehavioral condition in children and adolescents is attention deficit hyperactivity disorder (ADHD). A person with ADHD suffers from chronic inattention, impulsive hyperactivity, or both, impairing daily functioning. It is linked to a high rate of morbidity and impairment, and it affects up to 5% of individuals worldwide. ADHD has become a contentious topic, with some physicians and specialists believing that the disorder is underdiagnosed, with many youngsters going undiagnosed.

ADHD affects about 4.4 % of the adult population in the United States, while this figure is likely to be underestimated because up to 85 % of children with ADHD are at risk of developing the illness as adults, and only 10.9 % of adults with ADHD receive treatment.

Because ADHD makes it difficult for children to pay attention in class, a student with untreated ADHD may miss out on important information. They may fall behind or receive poor grades as a result. Furthermore, ADHD can be a financial burden, as it is linked to higher healthcare expenditures for persons of all ages with the illness, which may represent the usage of social services, educational resources, treatment, and lost productivity by family members owing to ADHD-related activities.

Raising ADHD awareness enhances the quality of life for those with the disorder and those who care about them. Because ADHD is so common, almost everyone knows someone who has it. ADHD stigma and beliefs are widespread, and these myths have an impact on people with ADHD's self-perception. These beliefs may also discourage people from seeking help or make them feel guilty and ashamed if they do. Suicide attempts are three times more common in people with ADHD than in the general population. When stigma and beliefs obstruct therapy, people's lives can be jeopardized.

Current ADHD diagnosis is primarily clinical, based on history, examination, and the doctor's subjective experience; according to a survey, this method has a sensitivity of 70-90 %. It can be backed up by neuropsychological tests but given the wide range of cognitive profiles seen in ADHD patients, these are more of a support role than a diagnostic one. Significantly, a large variety of normal cognitive profiles with various strengths and weaknesses in these areas manifest with disturbed attention, impulsivity, and dysexecutive disorders, confounding the differential diagnosis. The high rate of misdiagnosis not only prevented many children with ADHD from receiving effective treatment in the early stages of the disorder, but also hampered their normal growth. As a result, a biomarker that might help reduce the inherent uncertainty in clinical diagnosis would be extremely beneficial.

In past years, a variety of procedures for diagnosing ADHD had been used. Following extensive research, it has been shown that Convolutional Neural Network, or CNNs for short, are a potentially useful tool for diagnosing. It takes the necessary elements and automatically learns them in order to analyze medical images. The primary part of the CNN model, the convolutional filters, learn and extract the

precise features needed for efficient medical image comprehension. They have excelled beyond the capabilities of human experts in a number of visual interpretation tasks.

To be able to carry out this diagnostic procedure some cost-effective and high temporal resolution, diagnostic tools such as Electroencephalography (EEG), Magneto-encephalography (MEG), Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Positron Emission Tomography (PET) are used in the diagnosis of preliminary analysis of brain functions. EEG measures the electric potentials of neuron activation in various brain areas. In this MS thesis project CNNs will be employed to classify EEG signals into normal and with ADHD categories.

Chapter 2

LITERATURE REVIEW

2.1 ADHD Disease

Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder characterized by impulsivity (both motor and non-motor) and executive dysfunction. Approximately 8.4% of children and 2.5 % of adults suffer from ADHD. ADHD is usually recognized in school-aged children when there is disruption in the classroom or difficulty with homework. It is more common in boys than it is in girls. [3, 4]

ADHD can have a negative impact on a person's life in a number of different ways, including difficulties in school, social skills problems, and tense parent-child relationships. Children who suffer the illness are more likely to experience long-term detrimental repercussions, such as decreased success in school and the workforce. Therefore, it is crucial to create techniques for the early identification of ADHD in children so that specialized and personalized therapy may be used to support healthy brain development. If these conditions are not identified, it will be challenging to manage the symptoms as an adult.

A lab test cannot be used to diagnose ADHD. The diagnosis procedure includes gathering information from parents, teachers, and others, filling out checklists, and receiving a medical evaluation (including a vision and hearing test) to rule out other medical conditions. The symptoms are not caused by the person's stubbornness or anger, nor are they caused by their failure to understand a task or orders.

The underlying neuropsychological pathology of ADHD is still unknown, according to the Diagnostic and Statistical Manual of Mental Disorders (DSM) and International Classification of Diseases (ICD) [5], which solely rely on the responses from parents and educators to specially designed questionnaires [6], subjective assessments might be countered in this diagnosis by objective tests like the EEG.

2.1.1 ADHD Types

There are three forms of ADHD, depending on which symptoms are the most prominent in the person:

- ✓ **Predominantly Inattentive Presentation:** Planning or completing a task, prioritizing chores, and following commands or discussions are all challenging for the person. The individual is easily distracted and forgets daily knowledge:
    - Doesn't quite follow procedures or makes unnecessary mistakes in school or in the workplace.
    - Has difficulty concentrating on activities and tasks, like classes, talks, or lengthy reading.
    - When addressed to, does not appear to listen.
    - Ignores directions and fails to perform schooling, chores, or job responsibilities.
    - Has issues with task and job organization.

- Loathes or resists tasks that involve persistent mental work, such as writing reports and filling out forms.
- Frequently misplaces items required for daily duties, such as school notes, books, keys, wallet, mobile phone, and spectacles.
- Is prone to being interrupted.

✓ **Predominantly Hyperactive-Impulsive Presentation:** The person makes a lot of gestures and talks a lot. Inactivity for long periods of time is difficult to attain (e.g., for a meal or while doing homework). Children that are younger are more likely to run, jump, or climb constantly. The individual is irritable and has difficulty with spontaneity. People that are impulsive regularly interrupt others, take stuff from others, and speak at inappropriate times. Individuals find it difficult to wait their turn or follow directions. Accidents and injuries are more common in impulsive people than in others. Some of the symptoms for this type are given below:

- Inability to remain seated
- Squirms in seat or fidgets with or taps hands or feet.
- Inappropriately runs or climbs.
- Constantly "on the move," as though propelled by a motor.
- He speaks excessively.
- Inability to play or engage in recreational activities in peace.
- Gives an answer before the question is ended
- Has trouble waiting for his or her turn, as in when waiting line.
- Halts or infringes on others' talks, games, or activities, or begins to use other people's property without a warrant.

- ✓ **Combined Presentation:** The person exhibits symptoms from both sorts of above-mentioned symptoms.

2.1.2 Causes of ADHD

ADHD is a neurological condition with a complex etiology. While it is clear that alterations in neurodevelopmental processes such as synaptogenesis, myelination, and neurogenesis are the primary cause, the causes of these neurodevelopmental changes are many. Preterm delivery, perinatal issues, nutrition during pregnancy, and heavy metal exposure have all been associated to them. Furthermore, there is strong evidence of a hereditary component to ADHD, and a mix of environmental and genetic factors cannot be ruled out.

Neurogenesis, synaptogenesis, myelination, neuronal and glial proliferation and migration are all causes of ADHD. Neuronal development is influenced from early embryogenesis, even though symptoms do not occur until infancy [7].

ADHD has a wide range of causes, including prenatal, perinatal, and genetic factors. Scientists are researching the causes and consequences of ADHD in order to better treat and reduce the likelihood of someone developing the disorder. Although the cause(s) and risk factors for ADHD are unknown, recent evidence showed that heredity plays a significant influence. Genetic variables have been linked to ADHD in recent studies. Aside from genetics, scientists are looking into other probable causes and consequences, such as: accidental brain damage, environmental hazards (such as lead) during pregnancy or at a young age, tobacco and alcohol consumption during pregnancy, premature childbirth and low weight at birth.

2.1.3 Treatments for ADHD

It is crucial to first understand the physiological foundation of ADHD, as with any other illness affecting brain function, in order to establish an acceptable treatment. ADHD, like other NDDs, is caused by aberrant brain development that affects neurogenesis, synaptogenesis, myelination, as well as neuronal and glial proliferation and migration. Even if symptoms do not appear until infancy, neuronal development is impaired from early embryogenesis onwards.

Preschoolers should get behavioral treatments, and older children may benefit as well. Effective behavioral therapies include parent training, classroom management, and peer interventions. Medication is recommended as a first-line therapy for older children. The most effective psychostimulants for treating core ADHD symptoms are methylphenidate and dextroamphetamine, which have relatively little side effects.

2.2 Artificial Intelligence

2.2.1 Artificial Intelligence in The Diagnosis of ADHD

Medical technologies: Artificial intelligence (AI) is fast expanding in the health industry to satisfy humanity's requirements. Deep learning algorithms are used in a variety of medical fields to analyze data from wearable watches, artificial pacemakers, and other monitoring devices that are implanted in a person's body. Thanks to augmented medical technologies, doctors can give their patients a more personalized treatment. Physicians who oppose the advancement of modern medicine, on the other hand, are opposed to it. Medical education needs to be updated to match current state-of-the-art technology in this field. AI features have aided many ailments, including epilepsy, hypoglycemia, and atrial fibrillation.

'Smart Homes': Smart home technology have contributed in the transformation of traditional homes into more personalized, well-connected hi-fi homes. With the use of technology that includes various types of sensors, wireless connections, and algorithms, we may build a home that better matches our needs. Smart homes may control a number of functions in the house, such as room temperature, lighting, garden irrigation depending on weather conditions, gap and window closing, and giving optimal burglary security. These residences contribute to energy saving by turning off lights when not in use, turning off the water flow when not in use, and so on.

Given the aforementioned, clinical assessments of ADHD call for precise and impartial diagnostic procedures. The disease and specific physical symptoms have been the subject of numerous studies over the years. In the context of our issue, numerous research employed this technique to acquire brain pictures in order to identify differences between people with ADHD and control patients. Mental problems are frequently diagnosed with MRI (Magnetic Resonance Imaging). Brain volumetric imaging [8] and iron level analysis [9] are two examples of initiatives to make diagnosis more objective. The active brain areas in older people have been studied using functional magnetic resonance imaging (fMRI) in certain ADHD research [10].

Many techniques have been applied in the past years for the diagnosis of ADHD using AI and some of these techniques are listed below:

To identify the ailment, the usage of the Support vector machines (SVM) algorithm, a well-known artificial intelligence technology has been employed. The key advantage of using SVM is that it helps to control the diagnostic problem's complexity. There

hasn't been any SVM-based study or development on ADHD. To boost overall identification accuracy, they use the GA-based Feature Selection Algorithm. Genetic algorithms are well-known for their ability to provide effective answers to extraordinarily complex problems. AI methods like SVM will become increasingly important in future ADHD detection applications.[11]

A patient-centered engagement effort includes Pepper humanoid robotic therapeutic applications for youngsters with attention deficit. The goal of this new therapeutic methodology is to support and improve therapeutic practice. Pepper comes with a tablet and two identical cameras. The tablet is used to get the patients to participate with the program, while the cameras capture their emotions in real time to identify their level of attention and any issues they may be facing. Interacting with the tablet is done through a series of exercises that are presented in the form of games. The workouts of the patients are examined and compared to the data obtained by the cameras. The combination of these data is utilized to develop therapeutic activity levels that are appropriate for the patient. Children show more interest when interacting with a humanoid robot, which psychologists say is due to the fact that robots have lower emotional richness than humans, making patients feel less fearful[12].

Ghassemi et al. employed the wavelet entropy, correlation dimension, and Lyapunov exponent in four different signal groups [13]. EEG data from 50 participants were recorded, and while performing the test, the person was told to complete the NoGo Continuous Performance Test (CPT).

The ADHD diagnostic tools created by Ahmadlou et al. [14] were built on the wavelet approach and neural networks. They used generalized synchronizations known as synchronization likelihoods and the Radial Basis Function (RBF) neural network classifier (SL). Closed-eye EEG data were submitted by 47 ADHD patients and 7 members of the control group.

To identify between participants with ADHD and those who did not, Tenev et al. employed an EEG power spectrum and a machine learning model in a different study [15]. Four different SVM types and voting situations were used to improve decision-making regarding the outcome (SVM).

Nonlinear characteristics were employed by Khoushnoud et al. to diagnose ADHD [16]. They employed Probabilistic Neural Network, Approximate Entropy (ApEn), and Largest Lyapunov Exponent (LLE).

A diagnosis made early in life is preferable, as was previously said. As a result, contemporary research has focused on the identification of ADHD in children. Mohammadi et al. employed the Lyapunov exponent, approximative entropy, and fractal dimension (FD) in 2016 [17]. In order to improve the classification outcomes, the best features were chosen as inputs to the multi-layer perceptron (MLP) neural network using the Double Input Symmetrical Relevance (DISR) and minimum Redundancy Maximum Relevance (mRMR) techniques. They used the EEG signals of 30 ADHD children and 30 neurologically healthy, age-matched children.

In another work, Riaz et al. [18] used a Support Vector Machine to classify information from the NeuroBureau ADHD-200 [19] database (SVM). These findings

are based on a type of functional magnetic resonance imaging (fMRI) termed resting state fMRI (rsfMRI or R-fMRI), which measures regional interactions that occur while a specific activity isn't being performed. The achieved classification accuracy was 0.818.

In an effort to detect ADHD patients, Oztoprak et al. [20] examined ERP (Event-related potentials) from EEG data collected during a test similar to the Stroop. With high resolution time-frequency domain features as input, they used an SVM classifier with recursive feature removal. With a test group of 10 participants, they were able to reach 100% accuracy, but the accuracy for the entire train set was 0.995.

Sun et al [21].'s use of radiomics for MRI-based ADHD diagnosis. In order to gather information that can be used for diagnosis, this method depends on the extraction of a substantial number of features from medical imaging. Cross-validation using random forests resulted in an accuracy of 0.737.

Additionally, relevant publications that promote machine learning over actimetry can be listed. In the Muoz et al. experiment [15], two accelerometers—one on each wrist and ankle—were used. Activity data from a group of young children (22 patients, 11 ADHD patients, and 11 healthy children) was gathered over the course of six school hours and evaluated using a CNN. While the ankle device's accuracy was 0.9375, the wrist device's accuracy was tested at 0.8750. The sensitivity levels were 0.6 and 0.8, respectively.

For the purpose of characterizing motion, Mahony et al. employed accelerometers and gyroscopes in [22]. With an accuracy, sensitivity, and specificity of 0.9565,

0.9444, and 0.95512, respectively, the best performing features could be identified using SVMs.

In adults and children who had been diagnosed with ADHD, Markovska-Simoska and Pop-Jordanova [23] evaluated the quantitative EEG. The authors assessed and analyzed the absolute and EEG powers as well as the theta/beta ratios for both children with ADHD and control individuals, and they made the assumption that the theta/beta ratio was larger in the ADHD patients. Additionally, they noted that the ADHD kids had greater absolute capabilities of the slow waves than the typical individuals (theta and delta).

Seung Hyun Lee et al. was successful in identifying ADHD with an accuracy level of 60 to 70% using a discrete wavelet transform and a self-organizing map clustering method. Accuracy was thought to be impacted by the use of various wavelet functions at various stages of feature extraction. [24]

Using functional magnetic resonance imaging (fMRI) brain pictures and magnetoencephalography (MEG) data, Sato et al. [25] and Monge et al. [26] calculated the complexity of the brain networks using the graph spectral entropy technique. The ailment that was discovered, according to sources, affects children in addition to those who have ADHD and other pathological and physiological diseases like epilepsy and sleep issues.

It is crucial to remember that earlier studies have achieved even better results by effectively differentiating between HC and ADHD using conventional machine learning algorithms applied to EEG data. For instance, (Mueller et al., 2010)

achieved a sensitivity and specificity of 91% in predicting the diagnosis of ADHD in a sample of 150 people using a combination of five response-inhibition ERP variables found by ICA (75 with ADHD). [33]

In a study with fewer participants (n = 36), Nazhvani et al. (2013) created an algorithm to determine the combination of time points at which the ERP amplitude maximized the accuracy of group discrimination, achieving an accuracy of 94.6% in differentiating people with ADHD from controls. [34]

Similar findings were made by Sadatnezhad et al. (2011), who reported diagnosis accuracy of 86.4% using a combination of spectral power and fractal features of EEG time series, with fractal features producing the best discrimination. [35]

Ahmadlou and Adeli (2010) produced a maximum accuracy of 95.6% based on the combination of theta band synchronization at the frontal and occipital electrodes and delta band synchronization at the electrode T5 and frontal electrodes. [36]

Relative theta readings obtained from nine frontal scalp electrodes were employed by (Abibullaev and An, 2012) to attain a maximum accuracy of 97%. But for all these models, a thorough search in the frequency or time domain was followed by manual identification of disease-specific EEG characteristics.

Chapter 3

ELECTROENCEPHALOGRAPHY

Millions of neurons make up the human brain, and these neurons play a crucial part in regulating how the body responds to internal and external motor and sensory events. Between the human body and brain, these neurons will serve as information transporters. Analyzing brain signals or images can help one understand the cognitive behavior of the brain. Human behavior can be represented in terms of motor and sensory states, such as clenching of the hands, lip movement, memory, and attention. These states are connected to a certain signal frequency, which aids in understanding the functionality of the complicated activity of the brain. The effective modality of electroencephalography (EEG) makes it possible to collect brain signals that are related to different states from the surface of the scalp.

An electroencephalogram (EEG), a non-invasive examination, records the electrical patterns in your brain. The exam helps identify conditions like seizures, epilepsy, head injuries, vertigo, headaches, brain tumors, and sleeping problems. Routine EEG is the term used to describe an EEG test that is conducted when the subject is awake for a brief period of time. During sleep, aberrant electrical activities that are difficult to detect with a standard EEG test may take place. It can be used to confirm brain death as well.

The billions of nerve cells that make up your brain produce exceedingly weak electrical signals that group together to form patterns called brain waves. During an EEG, wires and small electrodes are attached to your head. A conductive gel is used to secure each electrode to the scalp. In some circumstances, a cap with electrodes attached to it may also be employed. Depending on the cap being utilized, different electrode counts are employed. Between 12 and 64 electrodes can be used to conduct the test. The EEG machine detects your brain waves, amplifies them, and records them as a wave pattern on graph paper or a computer screen.

The American Clinical Neurophysiology Society suggested the 10-20 and 10-10 criteria for scalp electrode placement [41]. The numbers represent the spacing between neighboring electrodes that were implanted on the skull. The relative distance between an electrode and the area of the skull underneath it, for example, is 10% or 20% for the 10-20 standard [41,42]. The electrode location starts with a letter, such as F (frontal), C (central), T (temporal), P (posterior), and O (occipital), to indicate positioning and the left or right side of the brain [43]. Figure 2 [42] displays the names and positions of each electrode in the 10-20 (black circles) and 10-10 (gray circles) systems. The 10-20 approach is suitable for both clinical and non-clinical studies as well as event-related potentials research (ERPs) [42]. The 10-10 approach is suitable for gathering more precise EEG data [41,43].

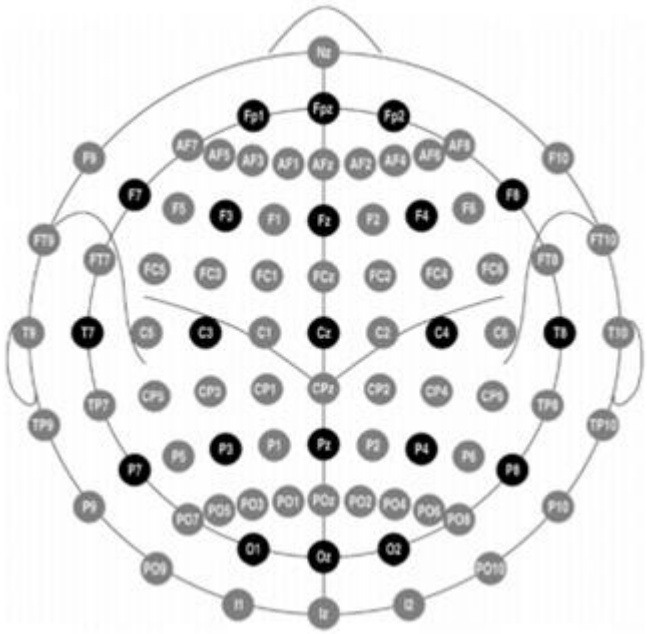

Figure 1. Electrode positions and labels in the 10 ± 20 system [42].

When the recording process begins, the electrical activity in the brain begins to be registered and stored into a storage device after all of the electrodes have been connected to the computer. The patient is made to feel at ease while the recording is being done. The patient might be instructed to perform easy math problems or stare at an image in order to stimulate more electrical activity in the brain during these recordings.

Different frequencies might be recorded at during the EEG test. The table below lists the various signal frequencies that range from 0.1 Hz to more than 100 Hz. These signals are often categorized into the following five groups: delta, theta, alpha, beta, and gamma.

Table 1: Brain wave frequencies with their characteristics

Type	Frequency (Hz)	Behavioral/ Psychological State	Neurotransmitter/ Hormone	Location
Delta	0-4	Deep rest, dreamless sleep	Human Growth Hormone, Melatonin	Frontally in adults, Posteriorly in children
Theta	4-8	Deeply relaxed	Serotonin, Acetylcholine, Anti-cortisol, Endorphins, Human Growth Hormone	Thalamic region
Alpha	8-13	Daydream, calm	Serotonin, Endorphins, Acetylcholine	Posterior regions
Beta	13-30	Alert, active thinking, anxiety, panic attack, focus, concentration	Adrenaline, Cortisol, Norepinephrine, Dopamine	Frontal and Parietal
Gamma	30-100	combination of two senses	Serotonin, Endorphins	Somatosensory cortex

According to [46], the EEG frequency bands from low to high frequencies are delta, theta, alpha, beta, and gamma. Delta waves appear during both deep sleep and waking (0.5-4Hz). Signals from the jaw and neck muscles that are artifacts are easily able to pollute delta waves. Theta waves (4–8Hz) are associated with intense focus

and unconscious activities. Early life and childhood are crucial periods for theta waves. There is no focus or attention present during alpha waves, which occur between 8 and 12 Hz. Beta waves are associated with thinking, concentration, and problem-solving (12–40Hz). Gamma waves are involved in conscious perception, working memory, and long-term memory (over 30 Hz). Muscle artifacts have a significant impact on gamma waves [47].

Evaluations are made after the EEG test results are complete, taking the patient's age and the medications used into consideration. These assessments can only be performed by a neurologist. The findings of the current test are compared to any prior EEG tests the patient has had. When the electrical activity in the brain is seen to be out of balance, it is assumed that there may be a brain-related health issue. The sickness brought on by this incompatible electrical activity is then identified by associated professionals by using various tests.

3.1 Why EEG?

The electroencephalogram (EEG), as opposed to NIRS or fMRI, is portable, non-invasive, easy to install, and has a fast time scale, making it the method of choice for imaging the human brain in the rapidly evolving Mobile Brain/Body Imaging paradigm.

An EEG can really pinpoint brain activity in contrast to MRI, which can only provide information about the physical composition of the brain. While MRI uses magnetic fields and radio waves to function, EEG uses electrical impulses. Since EEG is less expensive than MRI, it is usually chosen. EEG is quick and reasonably priced. An MRI costs a lot to buy, and it also has hefty maintenance expenses.

More than 90 years ago, EEG was initially used to examine ADHD [1]. Theta power activity rises whereas beta power sharply declines in the ADHD EEG signal, according to a 1991 article by Lubar [2]. He included these characteristics when developing a biofeedback system for therapy assessment and a diagnostic system for ADHD.

Because there are so many neurons producing currents, we may utilize electrodes put in a subject's head to measure some of the electric potentials that reach the scalp surface. German psychiatrist Hans Berger first employed electroencephalography (EEG), a neuroimaging technique, on humans in 1929. More recently, EEG was developed (BRONZINO, 1970). It records the evolution of the electric potential over time to produce a time series. The original purpose of EEG was to be a neurological and psychiatric diagnostic tool with a wide range of applications in the health and medical fields. These fields can be utilized to carry out surgeries, find abnormalities in neurophysiology, and advance neuroscience and psychology. EEG is a good option for use as a computer interface since it is non-invasive, transportable, and reasonably priced—especially on a large scale. Over some of those other methods, EEG has some advantages, such as:

- Compared to the bulk of other neuroimaging techniques, EEG is more forgiving of subject mobility. Movement artifacts in EEG data can even be eliminated using certain procedures.
- The ease of use of EEG makes it possible to track how the brain evolves throughout different stages of life. Analyzing EEG sleep data can reveal important details about the timing of brain development, including assessing adolescent brain maturity.

- There is more understanding of the signal that is obtained in EEG compared to other study approaches, such as the BOLD response in MRI.
- ERP studies can be conducted with far simpler paradigms than IE block-design fMRI experiments.
- Compared to the bulk of other neuroimaging techniques, EEG is more forgiving of subject mobility.
- A less intrusive alternative to electrocorticography, which places electrodes on the surface of the brain.

Wavelet transformation is typically used as a preprocess on EEG data before being fed into neural networks. This approach becomes a new favorite in recent research of EEG analysis using deep learning, up till the boom of deep learning and CNN. To achieve competitive accuracies on the dataset with clipped training, deep CNN has demonstrated a higher decoding performance.

The use of EEG was made possible by the fact that it is an excellent candidate for use in therapeutic settings, provided that current efforts to find other reasons of ADHD heterogeneity are successful through multivariate analyses and enhanced research on EEG signal generators.

Chapter 4

METHODOLOGY

Deep learning is a machine learning subfield inspired by artificial neural networks, which are themselves inspired by biological neural networks.

The convolutional network, often known as CNN or ConvNet, is one type of deep neural network. It's a feed-forward artificial neural network that's deep. Feed-forward neural networks are also known as multi-layer perceptron (MLP) models, which are the most common deep learning models. The models are known as "feed-forward" because data flows directly through them. There are no feedback links where the model's outputs are supplied back into it.

The biological visual brain has influenced CNNs in particular. Small clusters of cells in the cortex are sensitive to certain portions of the visual field. A fascinating experiment conducted by Hubel and Wiesel in 1962 developed on this concept. The researchers demonstrated that some individual neurons in the brain only activated or fired in the presence of edges of a specific orientation, such as vertical or horizontal edges, in this experiment. Some neurons activated when shown vertical sides, whereas others fired when shown a horizontal edge. Hubel and Wiesel discovered that all of these neurons were arranged in a columnar pattern and could produce visual perception when combined.[44]

Neural networks (NNs) perform well in tasks requiring supervised and unsupervised classification. However, NNs have one disadvantage—they could end up at local minima. The fact that there are numerous tuning factors for NNs, such as the number of neurons and learning rate, is another drawback. The number of parameters in a neural network increase quickly as more layers are added. In a neural network, the number of parameters and weights are generally equal. There will be a direct relationship between the number of layers and the number of neurons in each layer. The link between the number of parameters and overfitting is as follows: the likelihood of overfitting increases as the number of parameters increases. As a result, model training may become computationally challenging (and sometimes not feasible). It might be really difficult to tune so many different things. CNNs speed up the process of adjusting these settings.

Computer vision is a rapidly developing field. One explanation for this is deep learning. Convolutional neural networks are frequently used in computer vision, hence the term convolutional neural network (abbreviated as CNN). CNN is used in a variety of computer vision applications, including face recognition, object recognition, image classification, self-driving cars, segmentation, and image categorization. It has basic neural network characteristics. Similar to neural networks, CNN features learnable weights and biases. They provided cutting-edge research and outperformed conventional computer vision.

4.1 CNN Model

CNNs are fully linked feed forward neural networks. CNNs excel at reducing the number of parameters without compromising model quality. Images have a high feature density (each pixel is a feature), which is ideal for the above-mentioned capabilities of CNNs. CNNs were developed with images in mind, but they have also become industry leaders in text processing. CNNs have been trained to identify object edges in any image.

To lessen dimensionality, a sliding window that is smaller than the input matrix is employed. When we think intuitively, only a small piece of the total vision is taken into account at once. This square patch's window moves to cover the full image from left to right and top to bottom. All of a CNN's layers are active convolutional filters that accomplish dimensionality reduction by scanning the entire feature matrix. CNN is a great network for categorizing and processing images as a consequence.

CNN surpasses feed-forward networks because it has features like parameter sharing and dimensionality reduction. By utilizing parameter sharing, CNN uses fewer computations and fewer parameters overall. The fundamental tenet is that knowledge gained in one aspect of the picture may be used to another. Because of the reduction in dimensionality, CNN requires the least amount of computing power. An overview of the CNN model is given below.

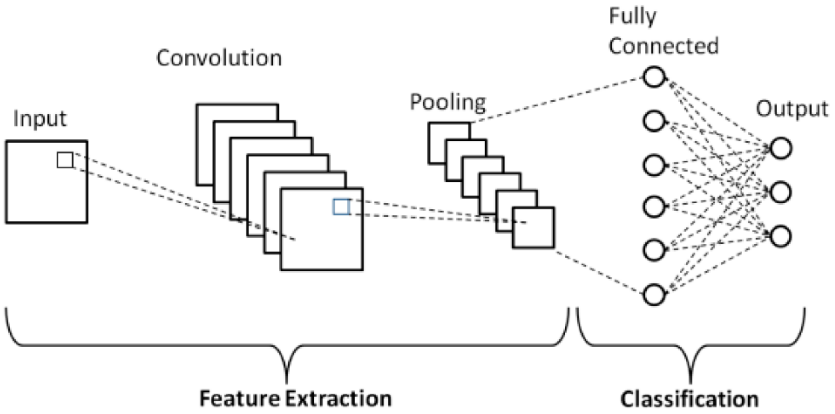

Figure 2: Schematic diagram of a basic convolutional neural network (CNN) architecture [40]

While utilizing the human eye to identify items may be straightforward, applying picture recognition algorithms in real-world situations is more difficult. The creation of picture identification software is difficult because there are so many options in the high-dimensional space of an image, such as color channels, forms, and different view angles. The classification of handwritten digits using Convolutional Neural Networks (CNN), which are specifically made to handle 2D forms, is a classic picture identification challenge.

While CNNs are similar to other neural networks, they use a series of convolutional layers, which increases their complexity. The convolution operation involves performing a mathematical operation on each region of the matrix, generating an activation map from the initial matrix, and slapping a filter (or kernel) matrix over another matrix while taking into account the padding number of additional rows and columns added to the initial input. CNNs are widely used in photo recognition tasks, which is primarily due to the properties of the convolution operation. By employing a filter that can recognize ever-more-complex patterns, the convolution process enables feature extraction to evolve spontaneously. The first convolution layer in a multi-layered design is capable of recognizing edges. The subsequent one can be

capable of object recognition, and so forth. The second one is capable of distinguishing between textures.[28]

The entire field of vision is covered by a partial overlap of the receptive fields of various regions. Similar to this, neurons in a convolutional layer of the CNN only receive input from a subarea of the previous layer, but these subareas are shared by several neurons. The outputs of several neurons in a cluster are combined in the following layer by pooling layers. Each neuron in a fully linked layer (such a multi-layer perceptron) receives information from every component of the layer below. In this way, CNNs perform exceptionally well in computer vision applications by automatically extracting features from images.

Convolution and pooling layers are used by CNN to extract features from the input image. For classification, extracted features are sent into flattening and fully-connected layers. A CNN Model consists of several different layers, the ones shown in figure 2 and here are their definitions:[31]

- **Convolutional layer**: A set of filters, commonly referred to as kernels, are used to construct convolutional layers from an input image. The convolutional layer's feature map is an illustration of the input image with the filters applied. Convolutional layers can be combined to create ever-more complex models that can extract intricate details from images.
- **Pooling layer**: Pooling layers are a subset of convolutional layers in deep learning. The spatial dimensions of the input are reduced by pooling layers, which facilitate processing and use less memory. Additionally, pooling shortens the list of training variables while accelerating the procedure. The two categories of pooling are maximum

pooling and average pooling. In max pooling, the maximum value from each feature map is used, whereas in average pooling, the average value is used. Pooling layers are frequently used after convolutional layers to reduce the size of the input before it is sent into a fully connected layer.

- **Fully connected layer**: One of the most fundamental layers in a convolutional neural network is one with fully-linked layers (CNN). As the name suggests, each neuron in a fully-linked layer is completely related to every other neuron in the layer below it. When using the features that the earlier layers learnt to make predictions, fully linked layers are typically used at the end of a CNN. For example, if we were using a CNN to identify images of animals, the final Fully connected layer might use the knowledge gained by the earlier layers to determine whether an image contains a dog, cat, bird, or another animal.

To sum up the structure of the CNN model, 3 by 3 by k matrices are used by the CNN for all of its filters (being k the number of channels in the input images to the convolutional layer). A "ceros" padding of size 1 is applied to all edges in the output image to prevent the convolutional process from reducing the size of the vertical and horizontal dimensions by 2 units. A normalization procedure is included before the non-linear ReLU layer and after each convolutional layer during CNN training to improve convergence. The final max polling layer, which lowers the output images' vertical and horizontal diameters by a factor of two, comes after the first two convolutional layers.

The output layer, which is based on the softmax layer, assigns probabilities for each image to each group (ADHD or generally developing youngster) and chooses the group with a greater probability. Each acceleration image has been assigned to a group for the hypothesis testing technique.

4.2 Different Types of CNN Architectures

A list of some of the most popular types of CNN architectures follows:

- **LeNet**: LeNet was proposed by LeCuN in 1998. It is well-known because, historically speaking, it was the first CNN to show cutting-edge performance on tasks requiring the identification of hand-held digits. Without being affected by slight rotations, distortions, changes in location, or scale, it can categorize digits. Five convolutional and pooling layers that alternate, followed by two fully connected layers, make up the feed-forward NN known as LeNet. It made use of the basic characteristics of the image, such as the distribution of feature motifs across the entire image and the association between adjacent pixels. Therefore, it is advantageous to use convolution with learnable parameters to extract similar properties from several locations with minimal parameters.[31]
- **GoogLeNet**: The intricacy and depth of the GoogLeNet model developed in [29] surpass all earlier CNN architectures. Additionally, it provides a brand-new module called "Inception," which is even more important because it merges filters of different sizes and shapes into a single new filter. Nine "Inception" layers total, two pooling layers, and two convolution layers make up GoogleNet. Each "Inception" layer consists of six convolutional layers and one pooling layer.
- **ResNet:** The paradigm for residual learning developed enables the layers to learn residual functions with respect to the inputs they receive rather than learning unreferenced functions. They were able to demonstrate that this

method is especially helpful for training deeper networks because residual networks are simpler to improve and achieve high accuracy. The high cost of analysis resulting from the numerous parameters is this network's biggest flaw. However, since the initial Fully-Connected layer controls the majority of the parameters, deleting it can drastically reduce the number of parameters without degrading performance. [32]

4.2.1 AlexNet as The Chosen Architecture

The 2012 ImageNet Large Scale Visual Recognition Challenge saw the 8-layer AlexNet architecture surpass other non-deep learning methods [30]. This success has reignited interest in CNNs in computer vision. AlexNet took first place in the competition with an error rate of 15.3% throughout its 8 layers, 5 of which are "convolutional layers" some of which are followed by max-pooling layers and 3 "fully connected" layers with a final 1000-way softmax and more than 60 million free parameters. The Softmax Function and the Cross Entropy Function are frequently related. In CNN, the Cross Entropy Function is utilized as the Loss Function after the Softmax Function has been applied to ensure the model's accuracy and to maximize the usefulness of our neural network. Furthermore, Alexnet implements Adam optimizer as its optimization function. An optimizer is a procedure or method that alters neural network properties like weights and learning rates. As a result, it aids in decreasing total loss and raising precision. An algorithm for gradient descent optimization is called adaptive moment estimation. When dealing with complex problems involving a lot of data or factors, the strategy is incredibly effective. It is effective and uses little memory. It combines the "gradient descent with momentum" algorithm and the "RMSP" algorithm, intuitively. The images below show a basic and a detailed AlexNet architecture.

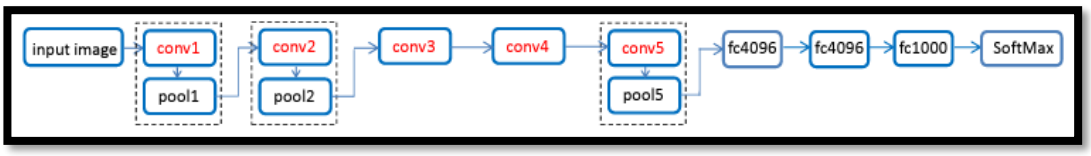

Figure 3: AlexNet Architecture [38]

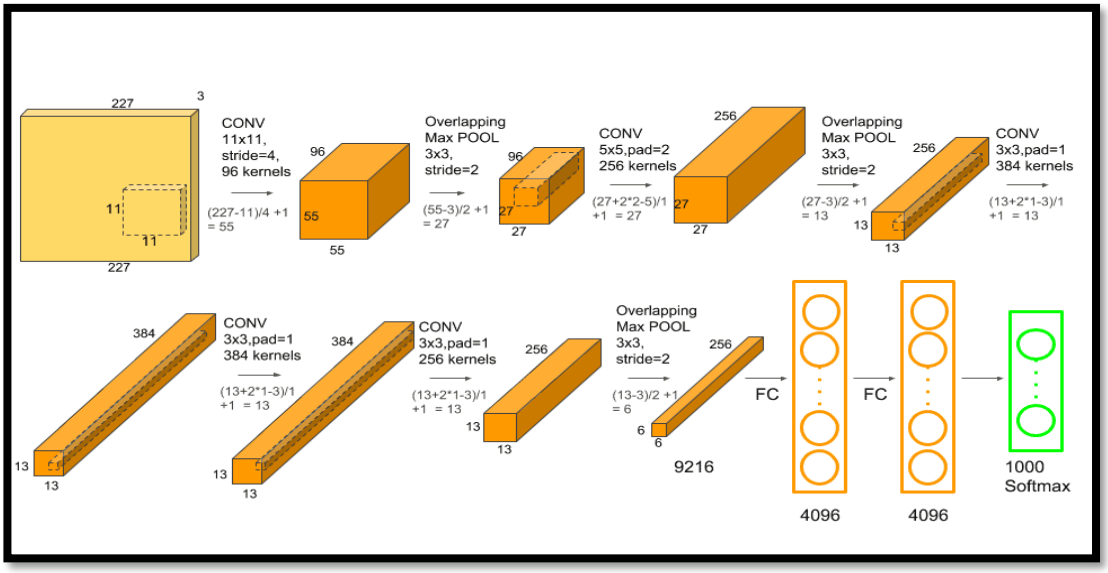

Figure 4: AlexNet detailed architecture [30]

AlexNet has a very similar structure to the LeNet architecture. However, there are important differences. AlexNet architecture is much deeper than LeNet, while AlexNet consists of 8 layers, LeNet consists of 5 layers. Overall, AlexNet has a large number of parameters. The input layer of the network can receive images, such as 227x227x3 images. This 3 denotes RGB. The layers and total number of parameters for AlexNet are shown in the table below. All of the parameters in the 5 Conv Layers + the 3 FC Layers together make up AlexNet's total amount of parameters. The total reaches to a staggering 62,378,344.

Table 2: AlexNet's layers and parameters

Layer Name	Explanation
Input	Color images of size 227x227x3.
Conv-1	The first convolutional layer consists of 96 kernels of size 11×11 applied with a stride of 4 and padding of 0.
MaxPool-1	The maxpool layer following Conv-1 consists of pooling size of 3×3 and stride 2.
Conv-2	The second conv layer consists of 256 kernels of size 5×5 applied with a stride of 1 and padding of 2.
MaxPool-2	The maxpool layer following Conv-2 consists of pooling size of 3×3 and a stride of 2.
Conv-3	The third conv layer consists of 384 kernels of size 3×3 applied with a stride of 1 and padding of 1.
Conv-4	The fourth conv layer has the same structure as the third conv layer. It consists of 384 kernels of size 3×3 applied with a stride of 1 and padding of 1.
Conv-5	The fifth conv layer consists of 256 kernels of size 3×3 applied with a stride of 1 and padding of 1.
MaxPool-3	The maxpool layer following Conv-5 consists of pooling size of 3×3 and a stride of 2.
FC-1	The first fully connected layer has 4096 neurons.
FC-2	The second fully connected layer has 4096 neurons.
FC-3	The third fully connected layer has 1000 neurons.

After the first two layers, conv1 and conv2, AlexNet applies local response normalization, which causes the most active neurons to inhibit neighboring neurons in adjacent feature maps. With this normalization, EEG task identification is successful.

AlexNet can show different actions in non-linear situations. Most standard neural networks used tanh or sigmoid. Tanh or sigmoid non-linearity had been frequently

utilized to train neurons within neural networks; this was the standard activation function for describing the internal neuron activity within CNNs.

AlexNet uses Rectified Linear Units, or simply ReLUs. ReLU is more advantageous than sigmoid. The output of the previous convolution layer is subjected to the transfer function operation ReLu. Positive values inside neurons are kept when ReLu is used, while negative values are clamped down to zero. Due to the fact that ReLu uses gradient descent optimization, which occurs more quickly than other non-linearity approaches, the training process can be sped up. The ReLu layer also introduces network non-linearity, which is a benefit. Similarly, the associativity of succeeding convolutions is eliminated.

The Alexnet architecture also has the benefit of using dropout to prevent overfitting. Overfitting is the remembering of the training data that was provided to the model. This issue can be summed up as training accuracy of 99% and testing accuracy of 75%. In this instance, it indicates that the model has memorized rather than learned the information provided for the train. Droupout has been used by AlexNet to stop this.

4.3 EEG Data for ADHD/Control Children Dataset

The study included 60 healthy controls and 61 children with ADHD (boys and girls, ages 7-12). An experienced psychiatrist used DSM-IV [37] criteria to diagnose the ADHD in the young patients, and Ritalin was recommended for up to six months. The kids in the control group did not have any reports of increased behaviors, seizures, or psychiatric disorders.

At 128 Hz sampling frequency, 19 channels (Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1, O2) were used to capture EEG based on the 10-20 standard. The earlobe electrodes A1 and A2 served as the controls.

Due to the fact that one of the weaknesses in children with ADHD is visual attention, the EEG recording approach was based on visual attention activities. The kids were given a variety of images depicting cartoon characters and told to count the characters. Children could readily see and count the characters because there were between 5 and 16 in each image, and the quantity of characters was picked at random. Following the child's response, each image was flashed instantaneously and continuously to provide stimulation during the signal recording. As a result, the child's performance throughout this cognitive visual task controlled the length of the EEG recording (i.e., response speed).

4.4 Experimental Setup and Preprocessing

4.4.1 Gramian Angular Summation Field (GASF)

Wang and Oates [39] proposed GASF for the aim of transforming timeseries signals into images. Before encoding, the input time series data must be standardized between [-1,1]. The time-series signal is then normalized or scaled to convert it from a Cartesian co-ordinate to a polar co-ordinate. This transform keeps the temporal information of the input signal. The signal is twisted in the transform domain. The next stage is to search for any temporal relationships between any two points in the polar coordinates. The [n,n]-dimensional Gramian matrix is produced by applying the trigonometric cosine function to this problem. The number of sample points for the EEG time period is indicated above by the letter n.

Let *T= {t1, t2,…,tn}* stand for a signal of n samples, which when rescaled has the range [-1,1] and may be written using the following form:

$$\bar{\mathbf{x}} = \frac{x_i - max(\mathbf{x}) + (x_i - min(\mathbf{x}))}{max(\mathbf{x}) - min(\mathbf{x})}$$

Next, we can encode the 0 to 1 scale time series we get as the angular cosine and the time stamp as the radius. The next step is to calculate the polar coordinates, which are the radius, r, from the time stamp t and the cosine angle, from the normalized amplitude values, as shown in Equation.

$$\begin{cases} \phi = \cos^{-1}(\bar{x}_i), & -1 \leq \bar{x}_i \leq 1, \bar{x}_i \in \bar{\mathbf{x}} \\ r = \frac{t_i}{N}, & t_i \in \mathbb{N} \end{cases}$$

The constant N serves as a regularization factor for the polar space span and is set to N = 1 in the equation above. Because the time series values incline toward the corresponding defined angle over time, the time series behavior is observed differently in the polar coordinate mapping. The preservation of temporal relations and a one-to-one mapping of the time series to the polar coordinate results are the two key aspects of the polar coordinate representation. The polar coordinates of the normalized time series in the [1, 1] interval are located at the angle limits [0,]. Due to the different informational concentrations provided, the GASF should be able to do any categorization jobs.

The Gram matrix known as Gramian Angular Summation Field is created by computing the temporal correlations between the adjacent points (i,j) and summing the angles. You can write this as,

$$GASF = [\cos(\Phi_i + \Phi_j)]$$

GASF has two fundamental features. The 0 to 1 range of the data must first be normalized. Having the ability to use cross items to convert GASF into normalized

time series data. The second is to maintain spatial coordinates while maintaining absolute time relations.

The time series is viewed as a one-dimensional metric space for each time step when transformed to the polar coordinate system. GAFs provide a lot of benefits. They maintain their time dependence as the position shifts from top left to bottom right. High-level features that the deep neural network learnt can be used to recreate the time series starting from the main diagonal. Because they are made up of nxn gramian matrices, GAFs are huge in comparison to time series. It is feasible to scale back.

It is possible to adapt this technique to generate an image from a chosen timeseries sample. In this experiment, each EEG epoch is sampled, producing a GASF image of. As a result of the angles of the data points being combined, we get the Gramian Angular Summation Field. The EEG signal is split into smaller temporal segments of samples in order to calculate GASF. The diverse EEG recordings from the aforementioned dataset result in a total of 25,200 photos. In order to generate data for effective DNN model training, several image transformations, including image rotation, shifting, and shearing, are also used.

Using GASF, it has been shown to outperform raw data. It is foreseen that recurrent neural networks can be used in the future.

4.4.2 Preprocessing Steps

- ✓ The experiment was carried out on 20 Individuals, 10 having ADHD and 10 healthy individuals.
- ✓ The fronto-parietal executive control network's main hubs (Fp1, Fp2, F3, Fz, F4, P3, and P4) were covered by EEG recordings from seven different places.
- ✓ The data were digitalized with 24-bit resolution at a sampling rate of 500 samples/second.
- ✓ Duration of the EEG signal is 3 minutes, it was split into 180 segments. Each segment is 500 milli seconds.
- ✓ Using GASF, those timeseries signals were converted into pictures.
- ✓ The signal capture, preprocessing, feature extraction, and classification are the four basic steps of the suggested work. EEG data sets with 7 channels were submitted by patients with ADHD and those without the condition for testing and validation. Each individual's EEG time-series data from various channels is offered in several EEG data sets as (excel).xlsx files. During the experimental phase, EEG time series data were input into MATLAB for each subject (.xlsx file) for each channel.
- ✓ A variety of libraries and frameworks, including PyTorch, Keras, and Scikit-learn, was used to train machine learning models. Keras trains the classification models for the GASF pictures using Tensorflow as a backend in a Google Colab notebook. Their creators claim the following about it: Keras is a Python-based deep learning API that is built on top of the TensorFlow machine learning framework. It was developed with the intention of promoting rapid experimentation. The ability to go swiftly from conception to conclusion is the key to performing effective research.

4.5 Learning and Generalization

The ability of artificial neural networks to adapt to new, previously unseen input that comes from the same distribution as when the model was learned is referred to as generalization. It refers to applying the knowledge gained throughout the learning process to a new scenario by referring to previously unseen test data, so combining the new experience with prior experiences that are comparable in one or more aspects. Pattern examples that represent the training database are used to train neural networks. A network's structure and parameters are adjusted throughout the learning phase so that it can respond appropriately to input signals. It corresponds to understanding the mechanism that the learning data were built on from a statistical standpoint.

When we train our own neural networks, we must be aware of a phenomenon known as neural network generalization. This essentially refers to how well our model can learn from provided data and apply what it has learned elsewhere. There will be some data that the neural network will train on, and some data that will be used to verify the neural network's performance. We can say that the neural network has generalized well on the given data if it performs well on data, it hasn't been trained on.

In this study, several methods for avoiding overfitting during CNN training were considered. In this thesis, the number of layers and feature maps were carefully designed to avoid overfitting while maintaining sufficient capacity for the network to solve the difficult ADHD classification problem. Data augmentation was then carried out using a dropout technique at the fully connected layers that contain the majority

of the data. As a result, the 3D CNN model was effectively trained and delivered classification accuracy that was at the cutting edge.

4.6 Data Partition

4.6.1 Training Set

A learning algorithm uses a training set, which is a collection of patterns and target vectors, to develop a classification model. A separate dataset termed the test set, which has the same probability distribution as the training set, is used in these domains in order to get the best generalization performance feasible. This thesis uses 80% of the aforementioned dataset as a training set. This is the quantity that CNN Models train on the most frequently. This percentage guarantees that the learning algorithm sees more than half (50%) of the data instances, which is necessary for effective generalization.

4.6.2 Validation Set

To prevent overfitting in any learning system, a validation set that is independent from the training and test sets is necessary. For instance, if the best classifier for the problem is sought, the training set is used to train the different candidate algorithms, the validation set is then used to compare their various performances and select the best algorithm, and finally the test set is used to determine the performance of the chosen algorithm. The validation set accomplishes a number of objectives, including serving as a training set for the algorithm validation yet being excluded from either the final testing or the low-level training. A straightforward way is to put aside a subset of the training set and use it as a validation set. It's called the holdout approach. Cross-validation is a technique that involves repeatedly partitioning the original training set into training and validation sets. The dataset can be divided into two sets and used for training-validation and then validation-training, or the k-fold

cross-validation method can be used to repeatedly choose a random subset of the dataset as a validation set.

4.6.3 Purpose of Cross-validation

One of the most popular strategies for resampling data to determine genuine model prediction error and fine-tune model parameters is cross-validation. In practice, ten-fold stratified cross-validation is frequently used.

In a learning process when the goal is to fit a model with many parameters on a given dataset, the induction approach tends to maximize these parameters as best as is practical on the provided dataset. Overfitting is the term for this situation. This is more likely to occur when there are few data observations or many model parameters. To solve this, the algorithm's ability to fit the learning model is evaluated using a different set of data. This is advantageous when we have to choose between competing algorithms. The least-squares algorithm is hence selected as the learning strategy.

4.6.4 K-Fold Cross-validation

The technique only has one parameter, k, which determines how many groups should be formed from a given data sample. K-fold cross-validation is a common name for the procedure as a result. For example, k=10 for 10-fold cross-validation can be used in place of k in the model's reference when a specific value for k is supplied.

The dataset is randomly split into k equal divisions in this sort of cross-validation. The remaining k-1 partitions are used for training, with a single partition set aside for cross-validation. This technique is repeated n times (folds). The average performance of the algorithm is then calculated, indicating the method's performance. Because each data item is used for training and validation, cross-validation offers an

advantage over other methods. Each observation is used n times during training and once during validation. The most usually used regiment [6] is n=10 in this thesis, but k is flexible. The picture below shows a 10-fold cross validation utilized in this thesis. K Fold splits all samples into folds (if, this is identical to the Leave One Out approach), which are equal in size (if possible). Folds are used to learn the prediction function, and the fold that is left out is utilized to test it.

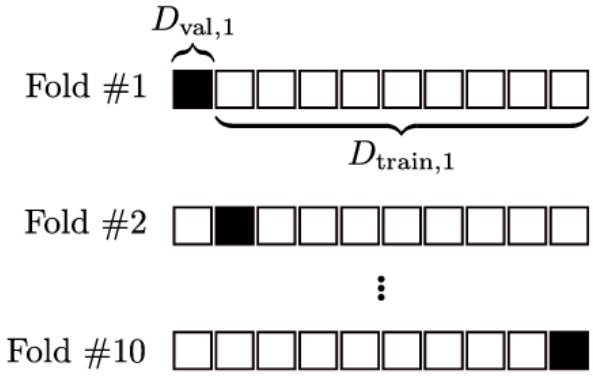

Figure 5: 10-fold cross-validation. [45]

Consider, for illustration, a learning set with n = 100 cases from the positive and negative classes, respectively, with n+ = 80 and n = 20. It is also probable that some validation sets will only contain positive examples if random sampling is used without stratification (or only negative cases). However, stratification ensures that each validation set in 10-fold cross-validation has roughly eight positive examples and two negative cases, mirroring the class ratio in the training set. The following forms the basis for stratified sampling. The population percentage can be accurately estimated using the sample proportion. The population percentage can be accurately estimated using the sample proportion. The class ratio in the learning set is the best indicator of the class ratio in the population since it provides a sample from the

population of interest. Therefore, data subsets that are used to assess the model should also take this class ratio into account to prevent a biased assessment.

The higher the k value given, the higher the success rate. However, each time the value of the k increase, the required processing load also increases.

Chapter 5

RESULTS AND DISCUSSIONS

The Custom CNN Model used to evaluate the transformed image. After utilizing GASF to acquire the photos, CNN was used to test the models for ADHD identification. Utilizing free and open-source deep learning frameworks like TensorFlow and Keras, custom CNN is created in the Python environment.

I used a 10-fold cross validation procedure to evaluate the model. In each fold, segmented samples were split into 80% for training and 20% for testing. In order to avoid overfitting the model, 20% of the segmented train data were used to test the model during training.

In this section, I present the experimental outcomes of the Deep-Learning (DL) technique used. Using the GASF approach, the EEG signals are first converted to RGB scale images which are shown in the figures below.

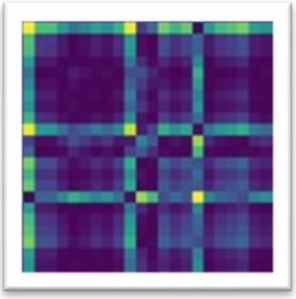

Figure 6: Representation of EEG signal using GASF for ADHD Patient

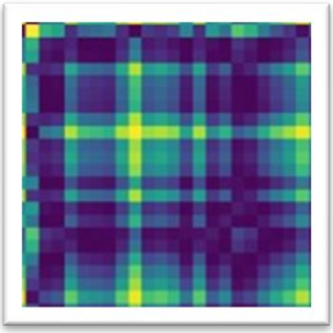

Figure 7: Representation of EEG signal using GASF for Healthy Individual

5.1 Evaluation Metrics

5.1.1 Confusion Matrix

Several metrics are used to assess how well medical image comprehension algorithms perform. The confusion matrix, also known as the error matrix, is the table used to visualize algorithm performance and establish various assessment criteria. It gives information about the different kinds of faults the classifier makes. It is a square matrix where the rows represent examples of the actual results, and the columns represent instances of the algorithm's projected results. The confusion Matrix has been computed various times with different various epoch, batch size and number of k-folds to find the highest model accuracy amongst them. An illustration of the confusion matrix is given below.

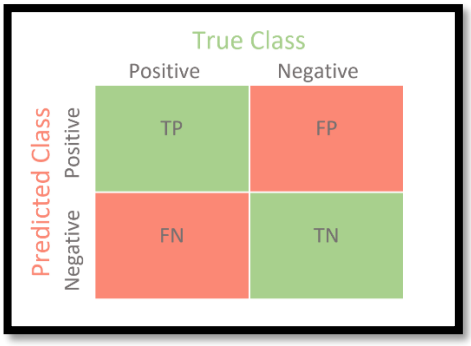

Figure 8: Confusion Matrix illustration

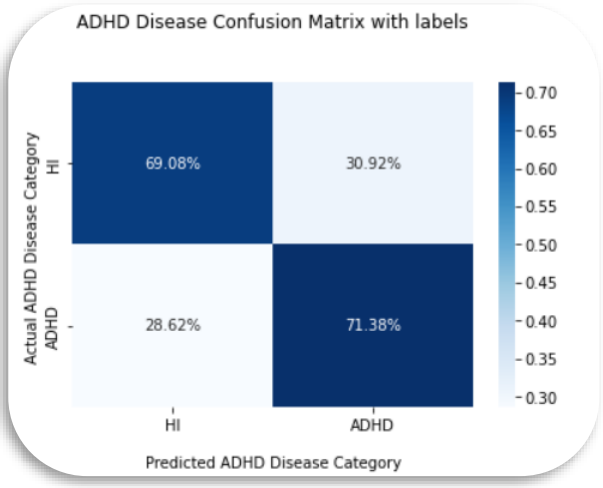

Figure 9: Confusion Matrix for ADHD Disease (20 Epoch, 256 Batch Size and 10 k-folds)

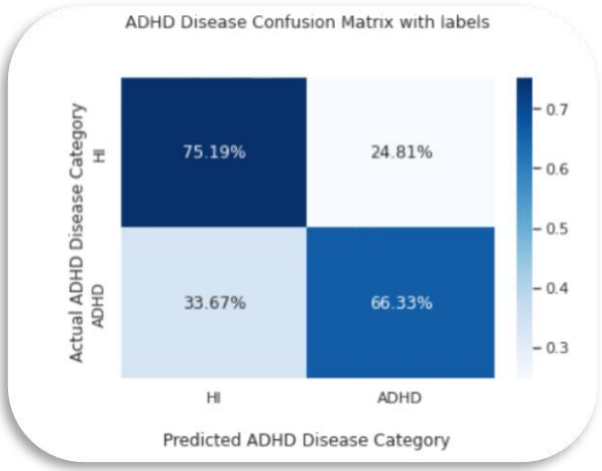

Figure 10: Confusion Matrix for ADHD Disease (30 Epoch, 128 Batch Size and 4 k-folds

5.1.2 Performance Metrics

Table 3: Performance Metrics

Performance Metric definition	Formula
Accuracy: Estimates the proportion of correctly predicted classes to all the samples that have been assessed.	Accuracy=TP+TN/TP+TN+FP+FN
Sensitivity or Recall: Used to calculate the proportion of correctly recognized positive patterns.	Sensitivity=TP/TP+FN
Specificity: Used to determine the percentage of negative patterns that are categorized correctly.	Specificity=TN/FP+TN

Where, respectively, T-ve, T+ve, F-ve, and F+ve stand for true-negative, true-positive, false-negative, and false-positive. The developed strategy is the optimum method for analyzing the given dataset when these values are close to unity, or the percentage value is close to 100.

According to this experiment, the TP, FP, FN and TN are redefined as follows:

- **True Positive (TP):** The model correctly identifies that the child has ADHD.
- **False Positive (FP):** The model predicts that the child has ADHD but the child is Normal.
- **False Negative (FN):** The model predicts the child does not have ADHD but he/she actually has.
- **True Negative (TN):** The model correctly identifies the child is Normal.

5.2 General Results

It could be seen how soon the network reached its highest level of accuracy and when it began to outperform the training samples by plotting the results of assessment metrics throughout the training stage.

Table 4: The average model accuracies for training and testing

State	Accuracy
Training	99%
Testing	71.7%

To examine the potential CNN models for ADHD identification, performance metrics like Accuracy, Sensitivity, and Specificity are computed. It should be noted that the CNN model generated an overall accuracy, sensitivity, and specificity of 0.717, 0.7461, and 0.8071 respectively.

Table 5: Performance Metrics Highest accuracies

Performance Metric	Percentage %
Accuracy	71.7%
Sensitivity	74.61%
Specificity	80.71%

5.2.1 General Results with K-FOLD

For each fold, I recorded loss and accuracy values at various training epochs, as well as loss and accuracy values during model validation in the table given below. In a maximum of 30 epochs and a batch size of 200, the model converged and achieved its highest level of accuracy.

Table 6: Folds' loss and accuracy values

Fold Number #	Loss value	Accuracy
1	1.80	69.28%
2	1.61	71.7%
3	1.58	69.16%
4	1.54	71.46%
5	1.59	69.95%
6	1.41	70.59%
7	1.61	70.31%
8	1.50	70.71%
9	1.87	68.60%
10	1.52	70.58%

The model was trained and tested using 10 K-FOLD Cross Validation. Those were done using the aforementioned channels. The obtained results are presented below:

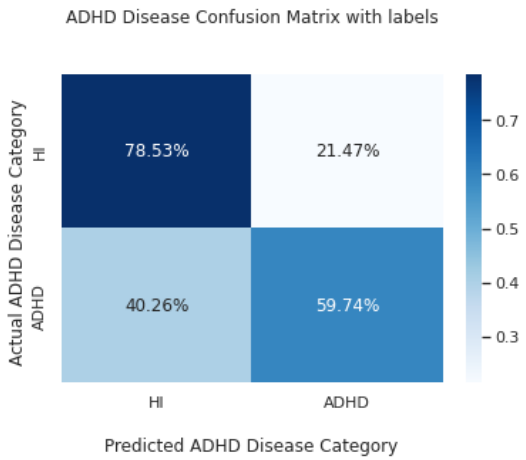

Figure 11: Fold 1 Confusion Matrix

- Accuracy = 0.6928260007927071
- Sensitivity = 0.5974235104669887

- Specificity = 0.7853239656518345

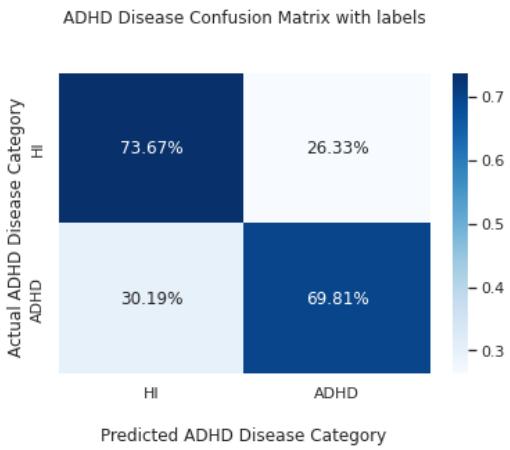

Figure 12: Fold 2 Confusion Matrix

- Accuracy = 0.7170035671819263
- Sensitivity = 0.6980544747081712
- Specificity = 0.7366720516962844

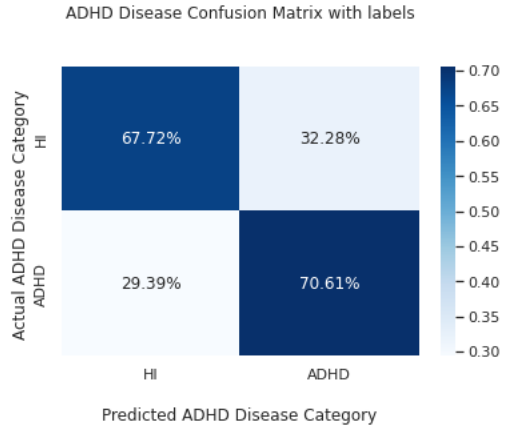

Figure 13: Fold 3 Confusion Matrix

- Accuracy = 0.6916369401506144
- Sensitivity = 0.7061159650516283
- Specificity = 0.6772151898734177

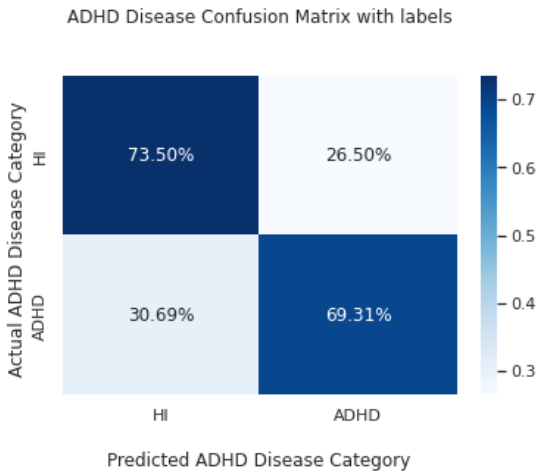

Figure 14: Fold 4 Confusion Matrix

- Accuracy = 0.7146254458977408
- Sensitivity = 0.6930612244897959
- Specificity = 0.7349768875192604

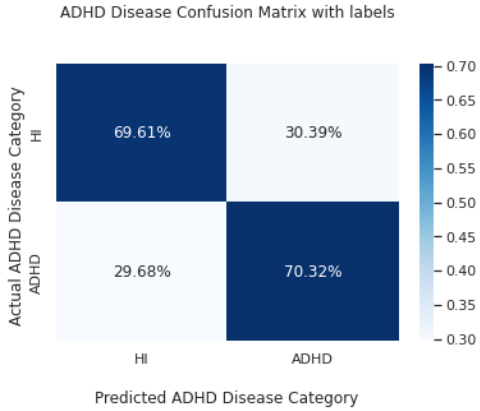

Figure 15: Fold 5 Confusion Matrix

- Accuracy = 0.6995640110978993
- Sensitivity = 0.7031630170316302
- Specificity = 0.6961240310077519

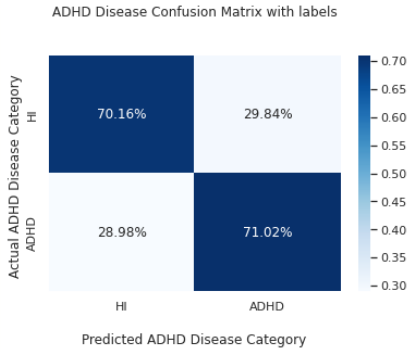

Figure 16: Fold 6 Confusion Matrix

- Accuracy = 0.7059056678557273
- Sensitivity = 0.7102137767220903
- Specificity = 0.7015873015873015

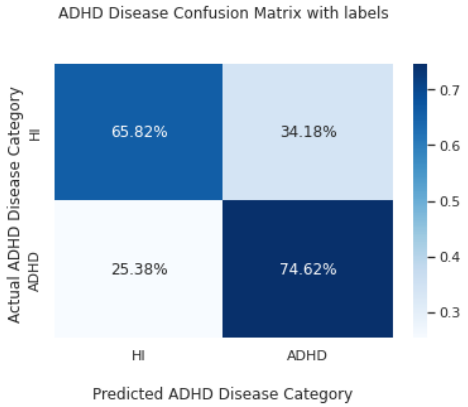

Figure 17: Fold 7 Confusion Matrix

- Accuracy = 0.7035275465715418
- Sensitivity = 0.7461538461538462
- Specificity = 0.6582174979558463

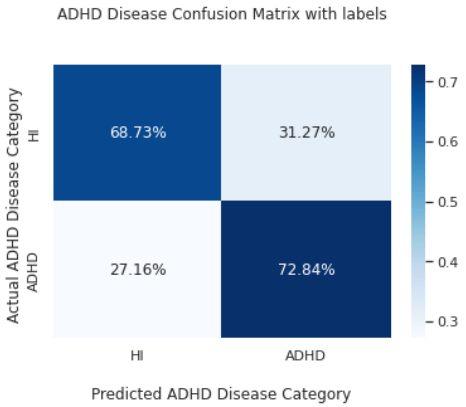

Figure 18: Fold 8 Confusion Matrix

- Accuracy = 0.7070947284978201
- Sensitivity = 0.7283950617283951
- Specificity = 0.6873088685015291

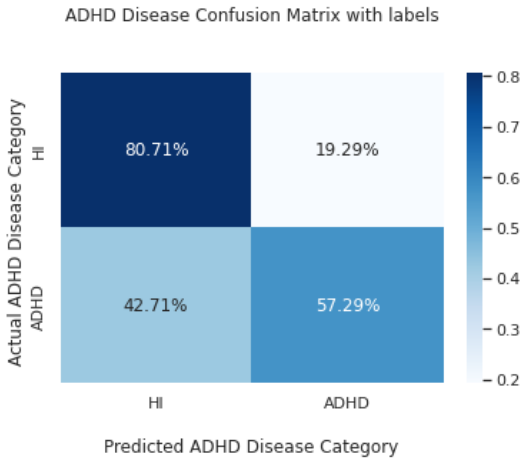

Figure 19: Fold 9 Confusion Matrix

- Accuracy = 0.6859635210150674
- Sensitivity = 0.5728527607361963
- Specificity = 0.8070607553366174

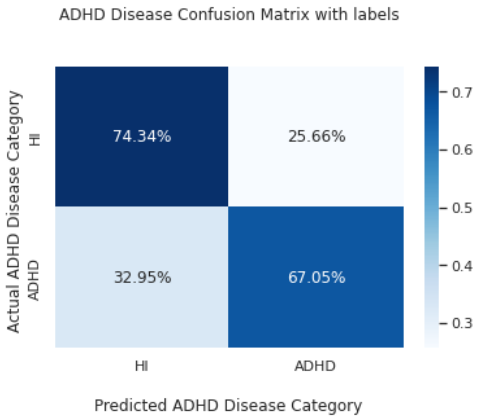

Figure 20: Fold 10 Confusion Matrix

- Accuracy = 0.7057890563045203
- Sensitivity = 0.6705069124423964
- Specificity = 0.7434426229508196

5.3 Discussion

Since deep learning methods for computer vision and image processing have become available, numerous researchers have used the simple CNN models in brain pictures to look for anomalies.

All the results obtained were compared with similar studies conducted in recent years. The following comparison tables represent the dataset used alongside the accuracies obtained.

Table 7: Comparison with other studies

Study	Year	Dataset	Features	Classifier	Accuracy
Majid et al. [48]	2020	31 ADHD children 30 healthy children	Frequency band separation Making RGB images	Deep CNN	98.48%
Allahverdy et al. [49]	2016	31 ADHD children 30 healthy children	Lyapunov exponent Katz fractal dimension Higuchi fractal dimension Sevcik fractal dimension	MLP NN	96.7%
Mohammadi et al. [17]	2016	31 ADHD children 30 healthy children	Approximate entropy Largest Lyapunov Exponent Katz fractal dimension Higuchi fractal dimension Petrosian Fractal Dimension	MLP NN	93.65%.
Chen et al. [5]	2019	50 ADHD children 51 healthy children	Mutual information Connectivity matrix	Deep CNN	94.67%

Chen et al. [50]	2019	50 ADHD children 57 healthy children	Filter Bank Common Spatial Patterns Gradient-weighted Class Activation Mapping	Deep CNN	90.29%
Dubreuil-Vall et al. [51]	2019	20 ADHD adult patients 20 healthy controls	Event-related potentials (ERP) during the Flanker Task ERP spectrograms using Wavelet	Deep CNN	88%
Zou et al. [52]	2017	ADHD-200	Histogram of Oriented gradient	SVM	62.57%
Ghiassian et al. [53]	2013	ADHD-200	Functional Connectivity Network	SVM	62.81%
Dey et al. [54]	2014	ADHD-200	fALFF, WC1 density	Multi modality 3D CNN with Concatenation	69.15%
This study	2022	EEG DATA FOR ADHD / CONTROL CHILDREN	GASF, Making RGB images from EEG signals	CNN	71.7%

Chapter 6

CONCLUSION AND RECOMMENDATION

6.1 Conclusion

Technology and computers will become much more complex than they are now as a result of artificial intelligence. Speech recognition technology will significantly progress, enabling it to communicate with humans in both text and speech using unstructured English. Expert system applications in clinical and administrative settings, for enhancing patient care and allocating financial, social, and other resources, will eventually have a bright future in all facets of health care. The problem of artificial intelligence producing computers that are smarter than humans, however, does not seem to have a solution.

A wide range of healthcare problems can be resolved using a variety of AI techniques. Despite earlier promises, medical AI technology has not been enthusiastically embraced. One explanation for this is how clinicians feel about using technology in the decision-making process. Surprisingly, accepting auto-analyzer biochemical data or magnetic resonance imaging images is not frowned upon. On the other hand, it is the duty of researchers studying this subject to show that these strategies are effective in real-world situations. Therefore, more randomized controlled trials are required to demonstrate the use of AI in medicine.

This study supports the notion that actimetry and deep learning algorithms can be combined to produce a quick and accurate technique for objectively diagnosing ADHD. Our findings demonstrate that our performance parameters (accuracy, sensitivity, and specificity for a case/control study) are comparable to or even better than those used in traditional techniques, which employ targeted trials in a controlled environment. Additional characteristics that are frequently considered when evaluating diagnostic procedure provide strong support for the suitability of our methods. As was already said, more research will be necessary to determine the traits found by the network's therapeutic importance.

The disorder of ADHD is untreatable. With medication and behavioral therapy, ADHD disease progression can be slowed down. The goal of this study was to use artificial intelligence to early detect ADHD in youngsters.

This study investigated the Gramian Angular Summation Field (GASF) method of translating EEG time series data to images for the diagnosis of ADHD. The work has benefited from the advantages of deep learning algorithms in picture classification tasks in order to classify EEG signals. The images created by the GASF method were fed into a CNN model for training and testing purposes. CNN was chosen as its integrated convolutional layer reduces the large dimensionality of images without losing any information. Therefore, CNNs are ideal for this use case.

The results examined were not as satisfactory as expected, those were 0.717, 0.7461, and 0.8071 for accuracy, sensitivity, and specificity respectively. They were lower than expected due to points that are mentioned below: The low accuracy was due to memory limitations due to a smaller dataset (imbalanced), therefore lack of enough

data for conducting such experiment. Secondly, due to the image method used which is in this case is GASF might not be applicable with such model. Finally, the CNN architecture chosen did not function as expected, therefore there is a need for repeating the process using different types of architectures to figure out which one provides best outcome. Furthermore, the relatively small dataset is one of the study's flaws, which when combined with the wide range of network parameters increases the risk of overfitting. Nevertheless, a number of regularization methods (other than early halting and dropout) could be able to solve this issue, and with larger datasets, improvements might be made.

Since our objective was to validate the notion that deep learning approaches can be valuable for the analysis of time frequency representations of EEG for the effective discrimination of ADHD rather than to outperform those models, I did not directly compare the performance of the current model with classical machine learning models in this thesis.

6.2 Recommendation for Future Work

Numerous revisions, testing, and experiments had to be delayed due to time constraints (i.e., the experiments with real data are usually very time consuming, requiring even days to finish a single run). Future research will concentrate on deeper exploration of specific mechanisms, innovative hypotheses to test novel approaches, or plain old curiosity. The following is a list of some concepts that could be used in future work:

- Using an alternative image method instead of GASF.
- Depending on their size or their unique significance in relation to the recognition process, the regions in the model and data images may be given

varying degrees of prominence. For example, this approach would make it easier to identify in extremely complicated situations the regions that must be located, the regions that occasionally emerge, and the regions that rarely do.

- Implementing another CNN architecture to attain better results
- In order to include some information on the variability among the many images and incorporate it to the attributes, the model's construction method might be altered as well. Rather than using one typical image (prototype), it may be based on different images. Regrettably, building a model from each image in the kind of photographs we used as real-world examples is a laborious process, thus no additional research in this area could be done.
- Adopting a larger dataset, since we have more information and hence more data, our estimate is more accurate. As our sample size increases, our estimate becomes more accurate, less uncertain, and more precise.
- Carrying out this experiment on adults also, not only on children.
- Employing other techniques instead of EEG. The encephalogram (EEG), magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), and positron emission tomography are some of these technological techniques (PET). The prime goal of all of these methods is to create accurate mental representations.
- Using Graphics processing unit (GPU) for graphics rendering. Multiple computations can be handled simultaneously by GPUs. As a result, machine learning processes can be dispersed, considerably accelerating them. With GPUs, you can increase the number of cores without sacrificing performance or power thanks to their lower resource requirements.

- The ability of GASF to detect the temporal dependence of ADHD in EEG has not been fully tested across a variety of sample windows. To test the GASF approach with different clinical data for the diagnosis of ADHD using deep learning techniques, more study is required.
- Future work should use larger training datasets and additional regularization strategies that more effectively prevent overfitting now that the system has been proven. As well as a deeper analysis of regularization and network building methods. Examples of this include deeper architectures, improved methods for data augmentation, and fresh approaches to data segmentation and standardization.

REFERENCES

[1] H. H. Jasper, P. Solomon, and C. Bradley, “Electroenchepalographic analyses of behavior problem children,” Am. J. Psychiatry, 1938, doi: 10.1176/ajp.95.3.641.

[2] J. F. Lubar, “Discourse on the development of EEG diagnostics and biofeedback for attention-deficit/hyperactivity disorders,” *Biofeedback Self. Regul.*, 1991, doi: 10.1007/BF01000016.

[3] M. L. Danielson, R. H. Bitsko, R. M. Ghandour, J. R. Holbrook, M . D. Kogan, and S. J. Blumberg, “Prevalence of Parent-Reported ADHD Diagnosis and Associated Treatment Among U.S. Children and Adolescents, 2016,” *J. Clin. Child Adolesc. Psychol.*, 2018, doi: 10.1080/15374416.2017.1417860.

[4] M. L. Danielson, R. H. Bitsko, R. M. Ghandour, J. R. Holbrook, M. D. Kogan, and S. J. Blumberg, “Prevalence of Parent-Reported ADHD Diagnosis and Associated Treatment Among U.S. Children and Adolescents, 2016,” *J. Clin. Child Adolesc. Psychol.*, 2018, doi: 10.1080/15374416.2017.1417860.

[5] H. Chen, Y. Song, and X. Li, “A deep learning framework for identifying children with ADHD using an EEG-based brain network,” *Neurocomputing*, 2019, doi: 10.1016/j.neucom.2019.04.058.

[6] A. Meysamie, M. D. Fard, and M. R. Mohammadi, “Prevalence of attention-deficit/hyperactivity disorder symptoms in preschool-aged Iranian children,” *Iran. J. Pediatr.*, 2011.

[7] L. Duan, L. Xu, Y. Liu, and J. Lee, “Cluster-based outlier detection,” *Ann. Oper. Res.*, 2009, doi: 10.1007/s10479-008-0371-9.

[8] L. Igual *et al.*, “A fully-automatic caudate nucleus segmentation of brain MRI: Application in volumetric analysis of pediatric attention-deficit/hyperactivity disorder,” *Biomed. Eng. Online*, 2011, doi: 10.1186/1475-925X-10-105.

[9] S. Cortese *et al.*, “Brain iron levels in attention-deficit/hyperactivity disorder: A pilot MRI study,” *World J. Biol. Psychiatry*, 2012, doi: 10.3109/15622975.2011.570376.

[10] D. Yu, “Additional Brain Functional Network in Adults with Attention-Deficit/Hyperactivity Disorder: A Phase Synchrony Analysis,” *PLoS One*, 2013, doi: 10.1371/journal.pone.0054516.

[11] J. Anuradha, Tisha, V. Ramachandran, K. V. Arulalan, and B. K. Tripathy, “Diagnosis of ADHD using SVM algorithm,” 2010. doi: 10.1145/1754288.1754317.

[12] F. Amato, M. Di Gregorio, C. Monaco, M. Sebillo, G. Tortora, and G. Vitiello, “Socially assistive robotics combined with artificial intelligence for ADHD,” 2021. doi: 10.1109/CCNC49032.2021.9369633.

[13] F. Amato, M. Di Gregorio, C. Monaco, M. Sebillo, G. Tortora, and G. Vitiello, “Socially assistive robotics combined with artificial intelligence for ADHD,” 2021. doi: 10.1109/CCNC49032.2021.9369633.

[14] F. Amato, M. Di Gregorio, C. Monaco, M. Sebillo, G. Tortora, and G. Vitiello, “Socially assistive robotics combined with artificial intelligence for ADHD,” 2021. doi: 10.1109/CCNC49032.2021.9369633.

[15] F. Amato, M. Di Gregorio, C. Monaco, M. Sebillo, G. Tortora, and G. Vitiello, “Socially assistive robotics combined with artificial intelligence for ADHD,” 2021. doi: 10.1109/CCNC49032.2021.9369633.

[16] S. Khoshnoud, M. Shamsi, and M. A. Nazari, “Non-linear EEG analysis in children with attention-deficit/ hyperactivity disorder during the rest condition,” 2016. doi: 10.1109/ICBME.2015.7404122.

[17] M. R. Mohammadi, A. Khaleghi, A. M. Nasrabadi, S. Rafieivand, M. Begol, and H. Zarafshan, “EEG classification of ADHD and normal children using non-linear features and neural network,” *Biomed. Eng. Lett.,* 2016, doi: 10.1007/s13534-016-0218-2.

[18] A. Riaz, M. Asad, E. Alonso, and G. Slabaugh, “Fusion of fMRI and non-imaging data for ADHD classification,” *Comput. Med. Imaging Graph.*, 2018, doi: 10.1016/j.compmedimag.2017.10.002.

[19] P. Bellec, C. Chu, F. Chouinard-Decorte, Y. Benhajali, D. S. Margulies, and R. C. Craddock, “The Neuro Bureau ADHD-200 Preprocessed repository,” *Neuroimage*, 2017, doi: 10.1016/j.neuroimage.2016.06.034.

[20] P. Bellec, C. Chu, F. Chouinard-Decorte, Y. Benhajali, D. S. Margulies, and R. C. Craddock, “The Neuro Bureau ADHD-200 Preprocessed repository,” *Neuroimage*, 2017, doi: 10.1016/j.neuroimage.2016.06.034.

[21] H. Sun et al., “Psychoradiologic utility of MR imaging for diagnosis of attention deficit hyperactivity disorder: A radiomics analysis,” *Radiology*, 2018, doi: 10.1148/radiol.2017170226.

[22] N. O’Mahony, B. Florentino-Liano, J. J. Carballo, E. Baca-García, and A. A. Rodríguez, “Objective diagnosis of ADHD using IMUs,” *Med. Eng. Phys.*, 2014, doi: 10.1016/j.medengphy.2014.02.023.

[23] S. Markovska-Simoska and N. Pop-Jordanova, “Quantitative EEG in Children and Adults with Attention Deficit Hyperactivity Disorder,” *Clin. EEG Neurosci.*, 2017, doi: 10.1177/1550059416643824.

[24] S. H. Lee, B. Abibullaev, W. S. Kang, Y. Shin, and J. An, “Analysis of attention deficit hyperactivity disorder in EEG using wavelet transform and self organizing maps,” 2010. doi: 10.1109/iccas.2010.5670255.

[25] J. R. Sato, D. Y. Takahashi, M. Q. Hoexter, K. B. Massirer, and A. Fujita, “Measuring network’s entropy in ADHD: A new approach to investigate neuropsychiatric disorders,” *Neuroimage*, 2013, doi: 10.1016/j.neuroimage.2013.03.035.

[26] J. R. Sato, D. Y. Takahashi, M. Q. Hoexter, K. B. Massirer, and A. Fujita, “Measuring network’s entropy in ADHD: A new approach to investigate neuropsychiatric disorders,” *Neuroimage*, 2013, doi: 10.1016/j.neuroimage.2013.03.035.

[27] Ali Motie Nasrabadi, Armin Allahverdy, Mehdi Samavati, Mohammad Reza Mohammadi, June 10, 2020, "EEG data for ADHD / Control children", IEEE Dataport, doi: https://dx.doi.org/10.21227/rzfh-zn36.

[28] L. Alzubaidi *et al.*, “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” *J. Big Data*, 2021, doi: 10.1186/s40537-021-00444-8.

[29] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” 2010. doi: 10.1109/cvpr.2009.5206848.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, 2017, doi: 10.1145/3065386.

[31] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, “A survey of the recent architectures of deep convolutional neural networks,” *Artif. Intell. Rev.*, 2020, doi: 10.1007/s10462-020-09825-6.

[32] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2016. doi: 10.1109/CVPR.2016.90.

[33] 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,” *Nonlinear Biomed. Phys.*, 2010, doi: 10.1186/1753-4631-4-s1-s1.

[34] A. D. Nazhvani, R. Boostani, S. Afrasiabi, and K. Sadatnezhad, “Classification of ADHD and BMD patients using visual evoked potential,” *Clin. Neurol. Neurosurg.*, 2013, doi: 10.1016/j.clineuro.2013.08.009.

[35] K. Sadatnezhad, R. Boostani, and A. Ghanizadeh, “Classification of BMD and ADHD patients using their EEG signals,” *Expert Syst. Appl.*, 2011, doi: 10.1016/j.eswa.2010.07.128.

[36] M. Ahmadlou and H. Adeli, “Wavelet-synchronization methodology: A new approach for EEG-based diagnosis of ADHD,” *Clin. EEG Neurosci.*, 2010, doi: 10.1177/155005941004100103.

[37] B. Abibullaev and J. An, “Decision support algorithm for diagnosis of ADHD using electroencephalograms,” *J. Med. Syst.*, 2012, doi: 10.1007/s10916-011-9742-x.

[38] W. Yu, K. Yang, Y. Bai, T. Xiao, H. Yao, and Y. Rui, “Visualizing and Comparing AlexNet and VGG using Deconvolutional Layers,” *Icml*, 2016.

[39] Z. Wang and T. Oates, “Imaging time-series to improve classification and imputation,” 2015.

[40] V. H. Phung and E. J. Rhee, “A deep learning approach for classification of cloud image patches on small datasets,” *J. Inf. Commun. Converg. Eng.*, 2018, doi: 10.6109/jicce.2018.16.3.173.

[41] J. N. Acharya, A. Hani, J. Cheek, P. Thirumala, and T. N. Tsuchida, “American Clinical Neurophysiology Society Guideline 2: Guidelines for Standard Electrode Position Nomenclature,” *J. Clin. Neurophysiol.*, 2016, doi: 10.1097/WNP.0000000000000316.

[42] R. Oostenveld and P. Praamstra, “The five percent electrode system for high-resolution EEG and ERP measurements,” *Clin. Neurophysiol.*, 2001, doi: 10.1016/S1388-2457(00)00527-7.

[43] P. A. Abhang, B. W. Gawali, and S. C. Mehrotra, *Introduction to EEG- and Speech-Based Emotion Recognition*. 2016. doi: 10.1016/C2015-0-01959-1.

[44] D. H. Hubel and T. N. Wiesel, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex,” *J. Physiol.*, 1962, doi: 10.1113/jphysiol.1962.sp006837.

[45] D. Berrar, “Cross-validation,” in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 2018. doi: 10.1016/B978-0-12-809633-8.20349-X.

[46] S. Sanei and J. A. Chambers, *EEG Signal Processing*. 2013. doi: 10.1002/9780470511923.

[47] A. S. Malik and H. U. Amin, *Designing EEG Experiments for Studying the Brain: Design Code and Example Datasets*. 2017.

[48] M. Moghaddari, M. Z. Lighvan, and S. Danishvar, “Diagnose ADHD disorder in children using convolutional neural network based on continuous mental task EEG,” *Comput. Methods Programs Biomed.*, 2020, doi: 10.1016/j.cmpb.2020.105738.

[49] A. Allahverdy, A. M. Nasrabadi, and M. R. Mohammadi, “Detecting ADHD children using symbolic dynamic of nonlinear features of EEG,” 2011.

[50] H. Chen, Y. Song, and X. Li, “Use of deep learning to detect personalized spatial-frequency abnormalities in EEGs of children with ADHD,” *J. Neural Eng.*, 2019, doi: 10.1088/1741-2552/ab3a0a.

[51] L. Dubreuil-Vall, G. Ruffini, and J. A. Camprodon, “Deep Learning Convolutional Neural Networks Discriminate Adult ADHD From Healthy Individuals on the Basis of Event-Related Spectral EEG,” *Front. Neurosci.*, 2020, doi: 10.3389/fnins.2020.00251.

[52] L. Zou, J. Zheng, C. Miao, M. J. McKeown, and Z. J. Wang, “3D CNN Based Automatic Diagnosis of Attention Deficit Hyperactivity Disorder Using Functional and Structural MRI,” *IEEE Access*, 2017, doi: 10.1109/ACCESS.2017.2762703.

[53] S. Ghiassian, R. Greiner, P. Jin, and M. R. G. Brown, “Learning to Classify Psychiatric Disorders based on fMR Images: Autism vs Healthy and ADHD vs Healthy,” *NIPS 2013 Work. Mach. Learn. Interpret. NeuroImaging*, 2013.

[54] S. Dey, R. Rao, and M. Shah, “Attributed graph distance measure for automatic detection of attention deficit hyperactive disordered subjects,” *Front. Neural Circuits*, 2014, doi: 10.3389/fncir.2014.00064.