EXPLAINABLE CNN-BASED ADHD DETECTION USING EEG DATA

By

AKASH BABU PEDAPAGA

A thesis submitted to the

Graduate School-Camden

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Computer Science

Written under the direction of

Dr. Sheik Rabiul Islam & Dr. Desmond Lun

and approved by

____

Dr. Desmond Lun, Ph.D.

____

Dr. Iman Dehzangi, Ph.D.

____

Dr. Michael A. Palis, Ph.D.

Camden, New Jersey

December 2024

THESIS ABSTRACT

AN EXPLAINABLE AI APPROACH FOR ADHD DETECTION USING CONVOLUTIONAL NEURAL NETWORKS ON EEG DATA

by AKASH BABU PEDAPAGA

Thesis Directors:
Dr. Sheik Rabiul Islam,
Dr. Desmond Lun

Attention Deficit Hyperactivity Disorder (ADHD) is a prevalent neurodevelopmental condition marked by persistent symptoms of inattention, hyperactivity, and impulsivity, significantly affecting individuals across all age groups globally. Accurate and timely diagnosis is critical for effective intervention, yet current diagnostic methods often rely on subjective clinical evaluations and behavioral assessments, which can be inconsistent and prone to bias. To address these challenges, this study introduces an innovative data-driven approach for the automated detection of ADHD using Electroencephalography (EEG) data, leveraging Convolutional Neural Network (CNN) models integrated with explainability techniques.

The proposed methodology employs advanced preprocessing techniques to extract meaningful features from raw EEG signals, capturing subtle neural activity patterns associated with ADHD. Utilizing a hybrid dataset comprising EEG recordings from both children and adults, the model demonstrates robust performance, achieving an accuracy of 98.91% on unseen test data. These results underscore the model's potential for precise and

reliable ADHD detection, offering a significant improvement over traditional diagnostic methods.

To ensure transparency and interpretability in clinical applications, two state-of-the-art explainability techniques—Local Interpretable Model-agnostic Explanations (LIME) and SHAPley Additive Explanations (SHAP)—were employed. LIME approximates the model's behavior for specific data instances, identifying influential features in individual predictions, while SHAP provides a global perspective by quantifying feature importance across the dataset. These techniques validated the relevance of specific EEG channels and features in distinguishing ADHD, revealing critical biomarkers and enhancing model interpretability.

This study establishes a comprehensive framework for automated ADHD detection, integrating deep learning with robust explainability methods to ensure accuracy and transparency. By bridging the gap between advanced machine learning techniques and clinical applicability, this work promotes objective, early, and reliable ADHD diagnosis. Beyond ADHD detection, the framework's adaptability suggests potential extensions to other neurodevelopmental disorders, highlighting its broader implications in AI-driven healthcare solutions.

ACKNOWLEDGEMENTS

I am deeply thankful to the Most High God I serve, who has been my strength, guide, and unwavering source of hope. He is the foundation of all that I am and all that I achieve.

I sincerely thank my research advisors, Dr. Sheik Rabiul Islam and Dr. Desmond Lun, for their invaluable guidance, encouragement, and support throughout this journey. My gratitude also goes to Dr. Iman Dehzangi and Dr. Michael A. Palis for their insightful feedback and contributions as members of my thesis committee. I am grateful to Dr. Sunil Shende for his mentorship, and to Sammed Sunil Admuthe for his contributions to the research paper.

To my parents, Mr. Vijaya Babu Pedapaga and Mrs. Elizabeth Rani Pedapaga, and my brother, Ajith Babu Pedapaga, thank you for your love and support, which have always been a source of strength. I also extend my heartfelt thanks to my aunt and uncle, Mrs. Roja Yellaula & Mr. Venkat Yellaula, whose encouragement has meant so much to me.

I am grateful to my Collingswood Church of Christ family for their care and kindness. Special thanks to Uncle Scott, Uncle Dave, Dr. Bill Ceraso & Family , Mr. Sean Sellers & Family, Mr. Tony Ceraso & the Ceraso Family and the McEntires for their wisdom and encouragement. I am especially thankful to my spiritual parents, Mr. and Mrs. Mark and Becky Finn, for their unwavering care, love, and inspiration.

This thesis is dedicated to all international students in the U.S. navigating the complexities of job searches and to unemployed neurodiverse individuals striving to overcome life’s challenges. May this work serve as a testament to perseverance, hope, and the strength to overcome adversity.

TABLE OF CONTENTS

THESIS ABSTRACT ii

ACKNOWLEDGEMENTS iv

TABLE OF CONTENTS v

LIST OF ILLUSTRATIONS vii

LIST OF TABLES viii

Chapter 1. Introduction to Neurodiversity and ADHD 1

1.1. Understanding Neurodiversity 1

1.1.1 Autism Spectrum Disorder (ASD) 1
1.1.2 Attention-Deficit/Hyperactivity Disorder (ADHD) 2
1.1.3 Neurotypical (NT) 2
1.1.4 Dyslexia 2
1.1.5 Dyscalculia 2
1.1.6 Tourette Syndrome (TS) 2

1.2 Understanding Neurodiversity in the Workplace 2

1.3 Understanding ADHD and Methods for Its Detection 4

Chapter 2. Background work 8

2.1 Importance of Neurodiversity in Cybersecurity 8

2.2 Skills and Strengths of Neurodivergent Professionals 10

2.2.1 Better Security Technology: 11
2.2.2 Improved Security Policies: 11
2.2.3 Creative Thinking 11
2.2.4 Analytical Abilities and Detail Orientation 12
2.2.5 Hyper-Focus, Passion, and Courage 12
2.2.6 The Benefits of Neurodivergent Talent in Cybersecurity 13
2.2.7 Innovative Thinking and Attention to Detail 13
2.2.8 Communication and Storytelling Abilities 13

2.3 Current Workforce Challenges and Misconceptions in Neurodiversity 14

2.4 Best Practices for Integrating Neurodiverse Talent in Cybersecurity 15

2.5 Creating an Inclusive and Supportive Workplace 17

2.6 Building and Supporting a Neurodiverse Cybersecurity Team 19

2.6.1 Fostering Neurodiversity in Cybersecurity 19
2.6.2 Leveraging Neurodiversity to Address the Cybersecurity Workforce Shortage 20

Chapter 3. Literature Review 21

3.1 Overview 21

Chapter 4. Dataset and Features 38

4.1 Dataset ***38***
4.1.1 Data Gathering 39
4.1.2. Data Preprocessing 40
4.1.3 Data Integration and Shuffling 41
4.2 Feature Extraction and Splitting ***43***

Chapter 5. Methodology 45
5.1 Model Architecture ***45***
5.2 Model Training and Evaluation ***45***
5.3 Prediction and Performance Metrics ***46***
5.4 Explainability with LIME ***46***
5.5 Explainability with SHAP ***47***

Chapter 6. Experimental Results 51
6.1 Receiver Operating Characteristic (ROC) Curve ***53***
6.2 Precision-Recall Curve ***53***
6.3 Training and Validation Loss/Accuracy ***54***
6.4 Aggregated Feature Importance ***55***

Chapter 7. Conclusion 60
7.1 Contributions ***60***
7.2 Future works ***62***
7.3 Final Remarks ***63***

References 64

LIST OF ILLUSTRATIONS

Figure 1: Flowchart illustrating the key steps in the ADHD detection methodology. 50

Figure 2: Receiver Operating Characteristic Curve 53

Figure 3: Precision-Recall Curve 54

Figure 4: Training and Validation Loss-Accuracy 54

Figure 5: Aggregated Feature Importance using LIME 55

Figure 6: Aggregated Feature Importance using SHAP 56

Figure 7: Sample Snapshot of EEG Electrode Placement on Human Head for Generating EEG Signals using EEGLAB from MATLAB toolbox 59

LIST OF TABLES

Table 1: Classification Report generated using the proposed model................................ 51

Table 2: Comparison of SHAP and LIME Importances with Snapshots 58

Chapter 1. Introduction to Neurodiversity and ADHD

1.1. Understanding Neurodiversity

Neurodiversity refers to the range of natural differences in how people think and process information. People with conditions like autism, ADHD, dyslexia, dyscalculia, and Tourette Syndrome are considered neurodivergent, meaning they think and interact with the world differently. “ND” is commonly used as an abbreviation for “neurodivergent” or “neurodiversity.”

Neurodiverse people process information and experience the world in unique ways, which are seen as differences rather than deficits. About 15%-20% of people globally identify as neurodivergent, while most others are considered "neurotypical.". Each type of neurodivergence represents distinct ways of thinking, learning, and interacting with the world, highlighting the importance of embracing diverse neurological perspectives. This chapter introduces the concept of neurodiversity, explores its relevance in both clinical and workplace contexts, and provides a foundation for understanding ADHD—a neurodevelopmental condition marked by inattention, hyperactivity, and impulsivity. Recent advances in machine learning (ML) and artificial intelligence (AI) offer promising avenues for addressing the challenges of ADHD diagnosis, particularly through non-invasive methods like EEG. Below, we explore key types of neurodivergence and their characteristics. These are some of the neurodivergence characteristics that were listed out by Jennifer E. Santiago as A Primer on Neurodiversity in Cybersecurity [1].

1.1.1 Autism Spectrum Disorder (ASD)

Autism affects how people perceive the world and communicate. People with autism may share common traits, like difficulty with planning and organizing, strict routines, sensitivity to sensory input, discomfort with eye contact, or difficulty reading body language.

1.1.2 Attention-Deficit/Hyperactivity Disorder (ADHD)

ADHD involves patterns of inattention (like trouble staying focused), hyperactivity (such as fidgeting or excessive talking), and impulsivity (like interrupting others).

1.1.3 Neurotypical (NT)

This term describes people whose brains function in ways that are considered typical, meaning they don't have neurodivergent traits.

1.1.4 Dyslexia

This learning difficulty affects language skills, especially reading. It can also impact pronunciation and other language abilities.

1.1.5 Dyscalculia

This condition affects a person's ability to perform basic math operations. Adults with dyscalculia may find working with numbers challenging and make more errors.

1.1.6 Tourette Syndrome (TS)

TS involves involuntary movements and sounds, known as tics. TS often occurs with other conditions like ADHD and OCD.

1.2 Understanding Neurodiversity in the Workplace

To begin, let's clarify essential terminology related to neurodiversity. "Neurotype" refers to a person's cognitive style—the way they perceive, process, and respond to information. "Neurotypical" is the term for the most common cognitive style, while "neurodivergent" includes variations like autism, ADHD, or dyslexia, which differ from the typical cognitive pattern. A “neurodiverse” group contains members with varying neurotypes, fostering diversity in cognitive processing and problem-solving.

Research done by McKinsey & Company [2] [3] shows that diverse workforces generally perform better, although studies specifically on neurodiversity in tech, and even more so in cybersecurity, are limited. Nonetheless, emerging findings suggest that neurodiversity offers advantages similar to other forms of diversity, enhancing both creativity and analytical rigor in teams.

Neurodiversity is the concept that human brains vary naturally in how they process information, think, and behave [4]. This variation includes differences such as attention deficit hyperactivity disorder (ADHD), dyslexia, learning disabilities, epilepsy, autism spectrum disorder (ASD), mental health conditions, and other neurological traits. Neurodiversity emphasizes that these differences are not deficits but rather unique ways of experiencing and interacting with the world [4].

The field of cybersecurity has a notable concentration of neurodiverse individuals [4]. Many roles in cybersecurity benefit from the particular cognitive strengths that neurodiverse people bring, which may explain why the industry attracts a higher proportion of these individuals. Mark Foudy [4] has said that Neurodiverse professionals often demonstrate strong skills in areas such as:

- Attention to Detail: They can focus deeply on specific tasks and catch subtleties others may overlook.
- Pattern Recognition: Many neurodiverse individuals excel at identifying patterns, a skill critical in detecting anomalies in cybersecurity.
- Logical Thinking: High logical reasoning ability helps in analyzing complex

systems, solving puzzles, and troubleshooting.

- Intense Focus: Certain neurodiverse traits enable prolonged concentration, which is especially valuable when working through complex security issues

Neurodiverse professionals bring a unique and valuable approach to cybersecurity. Their abilities in deep analysis, creative problem-solving, and innovative thinking make them key contributors to cybersecurity teams. By approaching challenges from different perspectives, they can devise solutions that others might not consider. As a result, neurodiverse individuals often play a crucial role in advancing cybersecurity practices and defending against evolving cyber threats.

Embracing neurodiversity in cybersecurity not only builds stronger teams but also fosters an inclusive environment where unique cognitive abilities are recognized as strengths [4].

1.3 Understanding ADHD and Methods for Its Detection

Attention-Deficit/Hyperactivity Disorder (ADHD) is one of the most common neurodevelopmental disorders, characterized by pervasive patterns of inattention, hyperactivity, and impulsivity that significantly impact an individual's cognitive, social, and emotional functioning. This disorder affects a diverse population across all ages and is often associated with a variety of comorbid conditions such as anxiety, depression, and learning disabilities, adding to the volume and complexity of data needed for accurate diagnosis. The early and accurate diagnosis of ADHD is critical for implementing effective interventions that can improve long-term outcomes; however, current diagnostic methods predominantly rely on subjective clinical assessments and behavioral observations, which are inconsistent and prone to bias. This highlights the urgent need for more objective, data-driven methodologies that can process vast amounts of information at high velocity and ensure veracity, thereby extracting valuable insights from diverse patient data to improve diagnostic accuracy and treatment effectiveness.

Recent advances in machine learning (ML) and artificial intelligence (AI) offer promising solutions for the detection of ADHD, particularly through the analysis of neuroimaging data. Electroencephalography (EEG) [5], with its high temporal resolution, provides a non-invasive and cost-effective method for capturing brain activity, making it a valuable tool for understanding the neural underpinnings of ADHD. In this context, Convolutional Neural Networks (CNNs) [6], a class of deep learning models particularly adept at handling high- dimensional data such as EEG signals, have shown significant potential in identifying patterns associated with ADHD.

Several recent studies have demonstrated the effectiveness of CNNs [6] in analyzing EEG data for ADHD detection. For instance, Taghi Beyglou [7] et al. proposed a novel approach combining CNNs [6] with classical machine learning classifiers such as Support Vector Machines (SVM), Random Forest (RF), and Logistic Regression (LR) to enhance diagnostic accuracy. Their method utilized raw EEG signals to train a CNN, which was followed by feature extraction to train classical ML classifiers, achieving accuracy rates like 91.16 ± 0.03% with CNN combined with LR, and 95.83% for unseen epochs. This integration of deep learning with traditional classifiers shows potential for practical clinical applications in ADHD detection.

Similarly, studies by He Chen [8] et al. and Dubreuil-Vall [9] et al. underscore the effectiveness of CNNs [6] in capturing neural signatures associated with ADHD. Chen [8] et al. used CNNs and Class Activation Maps (CAMs) to detect spatial-frequency abnormalities in children’s EEGs, achieving 90.29±0.58% accuracy. Dubreuil-Vall [9] et al. explored the capability of CNNs to distinguish adults with ADHD from healthy controls using event-related spectral EEG data, reaching 88 ± 1.12% accuracy, outperforming other models like Recurrent Neural Networks (RNNs). These findings highlight the value of deep learning in developing diagnostic tools for neurodevelopmental disorders.

Other studies have also made significant contributions to this field. Dahiru Tanko [10] et al. proposed EPSPatNet86, a specialized neural network architecture that achieved high accuracies of 97.19% in cross-validation and 87.60% in subject-wise validations, emphasizing the importance of customized network designs for enhancing diagnostic performance. Meanwhile, Ahire [11] et al. reviewed various EEG data classification techniques, focusing on the progress and challenges in early ADHD detection and the need for explainable AI models that clinicians can trust.

Building on these developments, this research introduces a comprehensive methodology that combines a CNN-based approach [6] for ADHD detection using EEG data with two advanced explainability techniques: Local Interpretable Model-agnostic Explanations (LIME) [12] and SHAPley [13] Additive explanations (SHAP) on a hybrid dataset that is generated by combining two open-source datasets : One dataset released by IEEE [14] and the second data released by Mendeley Data [15]. While the CNN model aims to achieve high diagnostic accuracy, the integration of LIME [12] and SHAP [13] addresses the critical need for transparency and interpretability in AI-driven diagnostic tools. LIME offers a localized understanding of the model’s behavior by explaining the influence of individual features on specific predictions, while SHAP provides a global perspective by quantifying the contribution of each feature across the entire dataset.

This dual approach enhances the model's reliability and interpretability, ensuring that its predictions can be trusted and understood by clinicians and researchers. By developing a framework that integrates high-performance deep learning approaches with robust explainability methods, our work aims to bridge the gap between machine learning and clinical practice, contributing to the early, objective, and reliable detection of ADHD , with applicability to other neuro-developmental disorders.

Chapter 2. Background work

2.1 Importance of Neurodiversity in Cybersecurity

The cybersecurity field has a large gap in available talent, with around 500,000–700,000 unfilled jobs [1]. Meanwhile, 60%–80% of adults with ASD are unemployed or underemployed. Given the importance of cybersecurity to national security, tapping into the skills of neurodivergent professionals is crucial. Although some neurodivergent individuals may communicate or socialize differently, companies like Microsoft, Google, and government agencies are starting to embrace their talents and address these gaps.

In cybersecurity, a diverse range of thinking styles is especially valuable. Matt Treadwell [16] has listed out the following strengths that Neurodivergent individuals bring, that support innovation and enhance security practices.

- Creative Problem Solving: Neurodivergent individuals often approach challenges from fresh perspectives, spotting potential vulnerabilities others might miss.
- Diverse Perspectives: Having team members with various ways of thinking allows for more effective security strategies, helping organizations stay agile and competitive.
- Pattern Recognition: Many neurodivergent individuals excel at noticing patterns and details, which is critical in identifying complex security threats.
- Focused Stamina: Neurodivergent individuals often demonstrate deep focus on tasks they enjoy, a strength in high-pressure cybersecurity environments.

While many organizations are focusing on diversity, equity, and inclusion (DEI) initiatives in terms of race and gender, top tech and STEM companies like Microsoft, SAP, and EY are now broadening their DEI efforts to increase neurodiversity [17]. Neurodiversity refers to the natural variation in human brains that leads to different ways of thinking, learning, and working. Initially, neurodiversity hiring programs focused mainly on autism, but many employers are now expanding their scope to include ADHD, dyslexia, and other cognitive variations that provide new perspectives and approaches.

Research [18] in psychology and neuroscience has identified specific strengths that neurodivergent individuals often bring to the workforce. For example, individuals with autism may excel in pattern recognition and systematic thinking—key skills for cybersecurity tasks such as monitoring networks and identifying security breaches. Many organizations have leveraged these talents in quality assurance and debugging roles within software development.

Similarly, those with ADHD and dyslexia often demonstrate high levels of creativity and an ability to draw connections across new ideas. Their innovative thinking and willingness to approach problems from unconventional angles are assets in the cybersecurity domain. Furthermore, many neurodivergent employees exhibit persistence, strong focus, and an exceptional commitment to problem-solving, making them invaluable to teams [18]. For these reasons, STEM companies frequently view neurodiversity as a source of competitive advantage.

As technology evolves with advances like artificial intelligence, the need for skilled cybersecurity professionals will only grow [17]. Addressing this demand will require a workforce that is not only creative and persistent but also skilled at detail-oriented, systematic thinking. These neurodivergent individuals are in today’s classrooms and can play a vital role in tomorrow’s cybersecurity landscape—if educators and cybersecurity companies make the effort to meet them where they are and recognize their potential.

Many individuals with neurodiverse conditions like ADHD, autism, and dyslexia offer fresh perspectives that can help tackle cybersecurity challenges. While technology plays a key role in cybersecurity, one of the greatest hurdles isn’t technical—it’s finding enough qualified people. For over a decade, both public and private sectors have struggled to recruit enough cybersecurity professionals, and the shortage continues to grow [17]. According to a 2022 ISC2 report [19], the U.S. alone has a cybersecurity workforce gap of 410,695, an increase of 9% from 2021. Globally, this gap is estimated at 3.4 million, with a 25% jump from the previous year. Approximately 70% of cybersecurity professionals surveyed reported that their organizations face a worker shortage.

2.2 Skills and Strengths of Neurodivergent Professionals

Holly Foxcroft [20] [21], head of neurodiversity research at the consulting firm Stott and May, describes neurodivergent people as having "spiky profiles," which means they often have standout strengths in certain areas. She explains that neurodivergent skills are often highly developed, particularly in problem-solving, creative thinking, and intense focus—essential abilities in cybersecurity.

Netflix tech blog [2] has listed out the following Complementary Strengths of Neurodivergent Professionals in Cybersecurity: Neurodivergent individuals often bring unique strengths that can be especially valuable in information security. While these strengths vary across individuals, many contribute meaningfully to cybersecurity roles. These abilities, though sometimes more prevalent in neurodivergent individuals, are also seen in neurotypical individuals. Here are some key strengths:

- Creative Thinking
- Visual-Spatial Reasoning

- Hyper-Focus, Passion, and Courage
- Innovative Thinking and Attention to Detail
- High Verbal Comprehension and Storytelling

Carpenter [20] [22] explains that hiring neurodivergent people aligns with cybersecurity's need for diverse problem-solving styles. Cybercriminals often bring unconventional thinking to their tactics, so defenders should, too. Neurodivergent workers, with their unique focus and creativity, help teams analyze data and find solutions that others might miss.

Matt Treadwell [16] claims that encouraging neurodiversity can directly improve a company's security posture. Key benefits that were listed out by Matt Treadwell [16] include:

2.2.1 Better Security Technology:

Neurodivergent individuals may identify usability issues faster, leading to more user-friendly and secure tools.

2.2.2 Improved Security Policies:

Neurodiverse team members are often skilled at identifying inconsistencies, helping build clear, straightforward policies that prevent errors and encourage compliance.

2.2.3 Creative Thinking

Creative thinkers excel in threat modeling, anticipating potential threats by imagining how attackers might breach systems [2]. This skill also enables them to find unconventional ways to test security controls, such as in red team exercises.

Paul Baird, a security officer at cybersecurity firm Qualys, adds that neurodivergent teams often respond more creatively to unexpected problems, as opposed to handling every issue in a standard way. This flexible approach is crucial for dealing with "unknown unknowns" or unforeseen challenges.

Perry Carpenter [20] [22], strategy leader at cybersecurity awareness firm KnowBe4 and someone on the autism spectrum, is pleased to see the cybersecurity community actively recruiting neurodivergent employees. He points out that while these individuals can bring unique strengths, it's important to avoid assuming that all neurodivergent people will excel in technical roles.

2.2.4 Analytical Abilities and Detail Orientation

Kieran Waite [20] [23], a consultant at the recruitment firm Hamilton Barnes, points out that individuals with autism or ADHD often have strong attention to detail and the ability to maintain focus, which are invaluable in cybersecurity roles that involve monitoring and responding to cyber threats.

High visual-spatial reasoning helps professionals quickly analyze complex attack paths in attack graphs, interpret malware code, and identify vulnerabilities [2]. This ability can be essential for understanding malware patterns or tracing intricate threat landscapes.

Example: At Netflix, Scott [2] used his visual-spatial skills to develop a methodology for mapping security risks against potential attack paths. This approach allowed the team to identify patterns across risks and focus on high-priority areas.

2.2.5 Hyper-Focus, Passion, and Courage

In fields like incident response, deep focus and resilience are vital. Neurodivergent individuals often excel at concentrating on specific tasks and making difficult, quick decisions under pressure, such as in forensics investigations.

Example: Melodie [2] from Netflix recounts using her passion and focus to dive deeply into bug bounty reports, ensuring no critical details were overlooked. Her dedication helped resolve vulnerabilities that improved platform security.

2.2.6 The Benefits of Neurodivergent Talent in Cybersecurity

Research in psychology and neuroscience highlights the unique skills that neurodivergent individuals can bring to cybersecurity [17]. For instance, individuals with autism often excel at detailed pattern recognition and systematic thinking—skills well-suited for tasks like security monitoring and breach detection. Similarly, those with ADHD or dyslexia may demonstrate strong abilities in idea generation and making connections between concepts, which can be invaluable for creative problem-solving. Many employers have found that the innovative and nonconformist thinking common among neurodivergent employees leads to fresh approaches and thorough, detail-oriented work. This ability to approach problems in new ways is increasingly recognized as a competitive advantage.

2.2.7 Innovative Thinking and Attention to Detail

These skills are particularly useful in vulnerability research, where noticing fine details and crafting new solutions is crucial [2]. Forensics work also benefits from meticulous attention to traces like altered timestamps or metadata.

Example: In leading data protection strategies, Scott [2] used innovative thinking to design a scalable, measurable program that met changing business needs and regulatory demands, supporting a sustainable cybersecurity framework.

2.2.8 Communication and Storytelling Abilities

Many neurodivergent professionals possess strong verbal comprehension abilities, allowing them to clearly convey complex security concepts [2]. This skill is vital in cybersecurity, where explaining the necessity of certain measures to stakeholders is crucial.

2.3 Current Workforce Challenges and Misconceptions in Neurodiversity

Addressing the cybersecurity skills gap is another motivation for hiring neurodivergent people [20]. Carpenter [22] highlights that broadening hiring practices isn't only ethically beneficial; it helps fill critical roles in cybersecurity. With an estimated shortage of 3.5 million cybersecurity professionals globally, finding talent from diverse backgrounds can bring new ideas to the field.

Devin Blewitt [20] [23], CIO at ITonlinelearning, emphasizes that other industries should also recognize the value of hiring neurodivergent talent. Holly Foxcroft [21] estimates that around 20% of the global population is neurodivergent, though the actual figure may be higher. She advocates for creating inclusive work environments and addressing biases, which can increase productivity and help attract and retain diverse talent.

Challenges That Still Exist: Pete Jarett [1] [24], a UK educator and advocate, explains that employers need to understand neurodivergent workers can excel by using different approaches. For example, dyslexia are due to the brain being wired differently—not wrongly. However, society often has narrow ideas of what “normal” should look like. Sadie Gauthier [1] [25], a neurodivergent writer, discusses how people sometimes view ability as a simple yes or no, overlooking her struggles on days when her health impacts her performance. Neurodivergent people are also frequently told, “You don’t look autistic,” highlighting misconceptions about invisible differences.

It’s essential to recognize that neurodiversity is just one part of a person’s identity. As the Autistic Self Advocacy Network (ASAN) notes [1], autistic individuals come from all backgrounds, ages, and identities—they can be of any race, religion, gender, or sexual orientation.

In his book Look Me in the Eye [26], John Elder Robison [1] describes his challenges with conversation, illustrating a common issue faced by neurodivergent individuals. Unlike a visible disability, conversational difficulties are often misunderstood, leading people to misjudge neurodivergent individuals as rude or indifferent when they may simply communicate differently.

2.4 Best Practices for Integrating Neurodiverse Talent in Cybersecurity

One challenge that companies face is actually finding neurodivergent talent [17]. Often, the traditional education system focuses on deficits rather than strengths, discouraging many neurodivergent learners from pursuing fields like STEM. This creates a scenario where brilliant problem-solvers may leave school feeling unqualified for professional careers in cybersecurity.

To attract neurodiverse talent, we need to rethink how we view neurodiversity in education. Shifting toward student-centered learning, where students' interests and talents guide their learning, can help neurodivergent learners discover their strengths [17]. By fostering confidence in their abilities, they can address areas of improvement without feeling limited by them. When learners are discouraged early on, they lose the opportunity to realize their potential in fields like cybersecurity.

Research [17] by Jodi Asbell-Clarke into successful neurodiversity hiring programs reveals that these companies adapt their recruitment, interview, and management processes in several key ways:

- Focus on skills and performance over traditional communication styles.
- Allow employees to work and express themselves in different ways, rather than adhering to a single cultural norm.

- Support clear communication through structured guidelines, meeting agendas, and detailed notes so nothing is missed or misunderstood.
- Encourage team-based work where each person contributes according to their strengths, allowing others to support areas where they may struggle.
- Many companies find that accommodations designed for neurodivergent employees actually benefit the entire workforce, as everyone appreciates clearer communication and support for diverse working styles.

Despite these advantages, organizations often struggle to attract enough neurodivergent talent [18]. Traditional education systems tend to focus on correcting perceived deficits in neurodivergent students rather than cultivating their unique strengths. As a result, many talented problem-solvers leave school doubting their own intellectual abilities, potentially missing out on promising careers in fields like cybersecurity. Employers can reduce the cybersecurity talent gap by actively recruiting neurodiverse candidates. Additionally, these inclusive practices stated by Michael Haynes [27] improve the work environment for everyone.

2.4.1 Sourcing

To attract neurodiverse candidates, ensure job postings are clear and straightforward, focusing on essential duties and avoiding unnecessary jargon. Terms like "social skills" or "interpersonal skills" should be used thoughtfully, as they may discourage qualified neurodiverse individuals. Including a statement about providing accommodations encourages more applicants to apply.

2.4.2 Recruiting

During recruitment, communicate each step of the interview process in advance, outlining clear expectations and timeframes. Since some neurodiverse candidates may struggle with conventional interviews, consider incorporating simulations or task-based activities to showcase relevant skills. When interviews are needed, choose a quiet space with minimal distractions to create a more comfortable environment.

2.5 Creating an Inclusive and Supportive Workplace

Encouraging neurodiversity can directly improve a company's security posture [16]. Matt Treadwell [16] has stated some key benefits that include:

2.5.1 Better Security Technology

Neurodivergent individuals may identify usability issues faster, leading to more user-friendly and secure tools.

2.5.2 Improved Security Policies

Neurodiverse team members are often skilled at identifying inconsistencies, helping build clear, straightforward policies that prevent errors and encourage compliance.

Building an inclusive workplace benefits everyone. Key steps include:

2.5.3 Flexible Environments

Options like quiet workspaces, clear expectations, and flexible schedules support neurodivergent employees' productivity.

2.5.4 Neurodiverse Hiring Practices

Simplifying job descriptions, interview formats, and providing clear instructions can make hiring more inclusive.

2.5.5 Building Allyship

Creating a supportive, psychologically safe environment lets neurodivergent employees feel valued and encourages open communication.

By embracing neurodiversity, cybersecurity teams can become more effective, innovative, and resilient, with strengths that make a genuine difference in security outcomes.

2.5.6 Retention

Once a neurodiverse employee joins your team, it's essential to create an environment where they can thrive. Michael Haynes [27] identified these retention strategies:

- Use plain language in communications, avoid jargon, and structure messages with bullet points to ensure clarity.
- Designate quiet work areas to help employees focus by reducing background noise.
- Streamline meetings, focusing on key points to avoid unnecessary information overload.
- Recognize that some employees may not feel comfortable speaking up in group settings. Share meeting agendas in advance and clarify what feedback is needed. Follow up individually after meetings to gather input.
- Offer neurodiversity awareness training to employees and managers. Integrating this training into existing programs fosters an inclusive culture across the organization.
- Establish mentorship or buddy systems to provide guidance with tasks, social cues, and document review.
- Provide additional tools as necessary, such as anti-glare screens, voice-to-text software, or screen readers, to help neurodiverse employees perform at their best.

2.6 Building and Supporting a Neurodiverse Cybersecurity Team

To cultivate a neurodiverse team successfully, Netflix tech blog [2] some recommended the following strategies:

- Inclusive Hiring Practices
- Use inclusive language in job descriptions.
- Offer clear and structured interview processes.
- Reduce distractions and provide flexible scheduling options.
- Supporting Neurodivergent Employees
- Respect different communication preferences, whether written or verbal.
- Provide clear guidelines on work expectations, including meeting etiquette and deadlines.
- Allow for flexibility in working hours and evaluate performance based on outcomes, not social expectations, like eye contact.
- Regularly seek feedback from neurodivergent employees to understand which adjustments could help them thrive.

2.6.1 Fostering Neurodiversity in Cybersecurity

Embracing neurodiversity in information security benefits organizations by fostering inclusive environments that promote innovation, productivity, and effective team collaboration [2]. Creating an inclusive setting for neurodivergent professionals enhances overall team performance, helps retain talent, and builds a dynamic culture of mutual respect.

2.6.2 Leveraging Neurodiversity to Address the Cybersecurity Workforce Shortage

A major challenge in cybersecurity today is not technological; it's finding enough skilled people to fill essential roles [18]. For over a decade, there has been a growing gap in the cybersecurity workforce across both public and private sectors, with demand consistently outstripping supply. According to the 2022 ISC2 report [19], the U.S. cybersecurity workforce faces a gap of 410,695 positions—a 9% increase from 2021. Globally, the shortage is even more severe, with 3.4 million unfilled roles, representing a 25% increase from 2021 to 2022. Nearly 70% of cybersecurity professionals in the ISC2 survey reported a staffing shortage within their organizations.

One solution proposed by ISC2 to close this gap is to embrace a more diverse talent pool, as cybersecurity requires a range of skills and perspectives to address complex problems.

Chapter 3. Literature Review

3.1 Overview

The detection of attention deficit hyperactivity disorder (ADHD) cases using a combination of Convolutional Neural Networks (CNN) and classical classifiers applied to raw Electroencephalogram (EEG) signals has garnered significant attention in recent research. Behrad Taghi Beyglou [28] et al., introduced a novel approach integrating CNN with classical machine learning (ML) models such as Support Vector Machines (SVM), Random Forest (RF), and Logistic Regression (LR). By training the CNN using raw EEG signals and subsequently utilizing the extracted feature maps to train classical ML classifiers, remarkable accuracies were achieved. The method demonstrated 86.33 ± 2.64% accuracy in a 5-fold cross-validation training set, with CNN combined with LR reaching 91.16 ± 0.03% accuracy. Notably, the approach achieved 95.83% accuracy in identifying ADHD in unseen epochs, showcasing its potential for practical clinical application.

He Chen [8] et al. proposed a pioneering method employing deep learning techniques to detect personalized spatial-frequency abnormalities in EEGs of children with ADHD. Leveraging CNNs alongside Class Activation Maps (CAMs) and Grad-CAM, the study achieved an impressive accuracy of 90.29% ± 0.58%. This method provides a precise spatial-frequency resolution for identifying abnormalities associated with ADHD, demonstrating promising results for personalized diagnostic approaches.

Similarly, Laura Dubreuil-Vall [9] et al., explored the efficacy of Deep Learning Convolutional Neural Networks (CNNs) in discriminating adult ADHD from healthy individuals based on event-related spectral EEG. The CNN architecture, consisting of four layers with a combination of filtering and pooling, was trained using stacked multi-channel

EEG time-frequency decompositions (spectrograms). The study reported an accuracy of 88 ± 1.12%, outperforming Recurrent Neural Networks (RNNs) and Shallow Neural Networks (SNNs). By incorporating techniques like Deepdream, the research underscored the potential of deep networks in analyzing EEG dynamics, offering insights into developing practical clinical biomarkers.

Furthermore, Dahiru Tanko [10] et al., introduced EPSPatNet86, a handcrafted model designed for detecting ADHD using EEG signals. The model incorporates various techniques including Directed Graph, 8-Pointed Star Pattern (EPSP), Tune able q Wavelet Transforms, Wavelet Packet Decomposition, Statistical Extractor, Iterative Chi2 Selector, and k-Nearest Neighbor (kNN) classifier. EPSPatNet86 exhibited high accuracies of 97.19% using 10-fold cross-validation and 87.60% using subject-wise validations, indicating its potential for widespread application in detecting EEG abnormalities associated with ADHD.

Nitin Ahire [11] et al., conducted a comprehensive review of EEG data classification techniques for ADHD detection using both machine learning and deep learning approaches. Highlighting the challenges in early ADHD detection and the potential of EEG as a neuroimaging technique, the review discussed various ML and AI strategies. Although the integration of ML/AI with EEG technologies in ADHD detection is in its nascent stages, the review suggests promising avenues for future research to develop reliable diagnostic tools for ADHD using EEG-based methodologies.

Hui Wen Loh [29] et al., proposed a deep neural network technique for automated detection of ADHD and Conduct Disorder (CD) using Electrocardiogram (ECG) signals. With the aim of creating an objective diagnostic model for ADHD/CD, the study utilized a 1-

dimensional CNN model trained on segmented ECG data. Achieving high classification accuracy, sensitivity, and precision, the proposed model demonstrated the potential for incorporating bio signal-based computer-aided diagnostic tools in clinical settings, contributing to enhanced healthcare management.

M Duda [30] et al., investigated the utility of machine learning models in behavioral distinction between Autism Spectrum Disorder (ASD) and ADHD. Given the subjective and time-intensive nature of current diagnosis methods, machine learning presents a promising avenue for quick and accurate assessment. By training and testing models on Social Responsiveness Scale score sheets, the study demonstrated the feasibility of accurately distinguishing between ASD and ADHD based on commonly measured behaviors, providing insights into developing more efficient diagnostic tools.

Smith K. Khare [31] et al., proposed an explainable model for ADHD detection using EEG signals, aiming to address the challenges in distinguishing between ADHD and healthy control EEG signals. Employing variational mode decomposition (VMD) and Hilbert transform (HT) for feature extraction, the model achieved high accuracy, sensitivity, specificity, and precision in detecting ADHD automatically. The model's interpretability was assessed using statistical analysis, offering insights into the EEG dynamics associated with ADHD and facilitating more accurate diagnosis.

Iqram Hussain [32] et al., investigated an EEG-based machine learning model for human activity recognition (HAR), emphasizing the potential of ML in recognizing everyday activities. By training ML models on EEG data collected from specific brain regions, the study demonstrated remarkable performance in distinguishing human activities. Incorporating Local Interpretable Model-Agnostic Explanations (LIME) enhanced interpretability, paving the way for improved activity monitoring in healthcare and clinical settings.

Deping Kuang [33] et al., proposed a modified deep learning method for identifying ADHD using ADHD-200 public dataset. By leveraging frequency features and deep learning techniques, the study achieved significant improvements in ADHD identification compared to existing approaches. However, the imbalance in datasets influenced classification results, highlighting the need for further research in addressing dataset biases for more robust ADHD diagnostic models.

Similarly, Francesco Carlo Morabito [34] et al., introduced an explainable Artificial Intelligence (XAI) approach for monitoring Mild Cognitive Impairment (MCI) using high-density electroencephalography (HD-EEG). By mapping EEG segments into channel frequency maps and training a CNN model, the study achieved high intra-subject classification performance, enabling accurate prediction of MCI to Alzheimer's Disease (AD) conversion. The model's interpretability was enhanced using a Grad-CAM approach, providing insights into the EEG biomarkers associated with MCI progression.

These studies collectively underscore the significant potential of machine learning, deep learning, and explainable AI techniques in enhancing diagnostic accuracy, interpretability, and clinical utility for ADHD detection and related neurodevelopmental disorders. The proposed methodology stands out by offering a more advanced approach, leveraging state-of-the-art algorithms and innovative feature extraction techniques to achieve superior accuracy and classification metrics compared to all other methods reviewed in the literature. This heightened performance not only provides more reliable diagnostic outcomes but also enhances the interpretability of complex data patterns, contributing to a deeper understanding of ADHD pathophysiology. Furthermore, the interdisciplinary nature of these research efforts, which integrates insights from neuroscience, psychology, and computer science, enables the development of robust diagnostic tools that are better tailored to clinical settings, ultimately leading to improved patient outcomes. Through these advancements, this body of work sets a new standard in ADHD diagnosis, fostering a paradigm shift towards more precise, explainable, and actionable clinical decision-making.

Joel Scanlan [35] et al. examined a structured framework to integrate neurodiverse individuals, particularly those with Autism Spectrum Disorder (ASD), into the cybersecurity workforce. The study highlights the alignment between neurodiverse cognitive traits—such as heightened attention to detail, pattern recognition, and sustained focus—and the specialized skill demands of cybersecurity. Addressing the sector's critical skill shortages, the authors propose a multi-phase model comprising talent identification, skill development, workplace deployment, and long-term evaluation.

Through gamified assessments, the study identified cognitive abilities in neurodiverse individuals that directly map to cybersecurity needs. These assessments also accounted for neurodiverse strengths in stress-minimized and engaging environments. Case studies, including the DXC Dandelion Program and Israel's Ro'im Rachok initiative, supported the effectiveness of structured accommodations in enabling neurodiverse individuals to excel in roles like security log analysis. Longitudinal evaluation of workplace integration underscored improvements in retention rates, performance, and satisfaction. The research concluded that leveraging neurodiverse talent can not only bridge the cybersecurity skills gap but also enhance diversity and resilience in the workforce.

The CREST report [36] explored the value of neurodiverse talent in addressing cybersecurity skill shortages. Highlighting conditions such as ADHD, autism, dyslexia, and dyspraxia, the study emphasized the unique strengths neurodiverse individuals bring to cybersecurity, including advanced pattern recognition, attention to detail, and methodical problem-solving. The research identified traditional hiring practices as a barrier, particularly their emphasis on social skills, which can exclude candidates with exceptional technical expertise.

The report advocated for clear job descriptions that distinguish between essential and desirable skills and called for technical-based interview processes to replace socially oriented evaluations. It also examined how sensory-friendly workspaces, structured task allocation, and flexible schedules could create environments where neurodiverse individuals thrive. Case studies, such as GCHQ's successful employment of individuals on the autism spectrum, demonstrated the exceptional performance of neurodiverse employees in highly analytical roles, like threat detection and anomaly analysis.

CREST concluded that integrating neurodiverse talent not only addresses immediate skill shortages but also enriches cybersecurity strategies through diverse cognitive approaches, enhancing the sector's capability to navigate complex security challenges.

Brenda K. Wiederhold [37] investigated the growing shortage of skilled professionals in cybersecurity and argued that neurodiverse individuals could play a critical role in addressing this gap. The paper underscores the essential role of cybersecurity in protecting networks and data amidst increasing threats and highlights the industry's inability to meet workforce demands. Wiederhold identifies neurodiverse individuals, including those with autism, ADHD, and dyslexia, as possessing cognitive strengths—such as problem-solving, pattern recognition, and attention to detail—that align closely with cybersecurity needs, particularly in areas like threat detection, anomaly identification, and network monitoring. The study critiques traditional hiring practices in cybersecurity, noting that conventional interviews often prioritize social and communication skills, which can disadvantage neurodiverse candidates despite their strong technical capabilities. Drawing on evidence from programs like MITRE's pilot and JP Morgan Chase's initiatives, the paper demonstrates that neurodiverse employees frequently outperform their neurotypical peers in tasks requiring high focus and analytical skills, with autistic employees at JP Morgan Chase showing up to 140% higher productivity in specific analytical roles.

Wiederhold advocates for redesigning recruitment processes to prioritize technical skills over interpersonal abilities and restructuring workplaces to better accommodate neurodiverse needs. Recommendations include creating precise job descriptions, implementing technical-based assessments, and introducing workplace accommodations like noise-cancelling headphones and flexible schedules to reduce sensory stress. Programs such as the DXC Dandelion Program serve as examples of how targeted initiatives can enable neurodiverse individuals to excel in cybersecurity roles when provided with supportive environments.

The paper concludes that neurodiverse talent not only helps fill the skills gap but also enriches cybersecurity with diverse approaches to problem-solving. By fostering innovation and inclusivity, organizations can address complex cybersecurity challenges more effectively, positioning neurodiversity as a strategic advantage in building robust and resilient defences.

Demek [38] et al. explored the talent shortage in accounting and information systems, proposing that individuals with Autism Spectrum Disorder (ASD) offer a viable solution to this challenge. The paper highlights the increasing reliance on technology in fields like accounting, cybersecurity, and data analytics, which has intensified the demand for specialized professionals. Despite this, many ASD individuals remain underemployed or unemployed, even though their cognitive strengths—such as attention to detail, logical reasoning, and pattern recognition—are well-suited to these fields.

The authors introduce a framework combining autism studies with decision-making literature to examine how ASD individuals process information differently from their neurotypical peers. The framework posits that ASD individuals are less prone to judgment errors and biases, such as heuristics and groupthink, due to their structured and analytical approach to problem-solving. This distinct cognitive style allows them to excel in roles requiring high precision, potentially reducing errors and enhancing efficiency in tasks like data analysis and risk assessment.

Case studies from neurodiversity programs at firms like Deloitte and EY provide practical examples of how ASD professionals contribute meaningfully in consulting and technology roles. These initiatives report benefits such as lower turnover rates and heightened innovation, underscoring the value of structured work environments and targeted recruitment strategies. ASD individuals, by avoiding cognitive shortcuts and offering unique insights, have demonstrated the potential to drive innovation and improve organizational outcomes.

The paper concludes by calling for interdisciplinary research to deepen understanding of how ASD individuals excel in these roles and to develop more inclusive hiring practices.

By addressing skill shortages and providing meaningful career opportunities for neurodiverse individuals, organizations can achieve mutual benefits, fostering diversity and innovation while strengthening technical capabilities in accounting, information systems, and cybersecurity.

Emmanuelle Walkowiak [39] investigates the intersection of digital transformation and neurodiversity, emphasizing the potential for autistic individuals to thrive in evolving workplaces shaped by automation and artificial intelligence (AI). The study highlights how digital advancements align with the cognitive strengths of autistic workers, such as high attention to detail, superior pattern recognition, and focused task execution. These attributes position neurodiverse individuals as valuable contributors to the workforce in the digital age, where such skills are increasingly in demand.

The paper introduces the theoretical concept of "productive complementarities," positing that neurodiversity and digital transformation are mutually reinforcing. For example, AI systems that require extensive data handling and monitoring can benefit from the precision and persistence of neurodiverse employees. Walkowiak further examines the role of inclusive workplace practices in facilitating this integration. Measures such as clearly defined job roles, sensory-friendly environments, and structured communication pathways are identified as key to creating supportive settings for neurodiverse individuals. The research demonstrates how companies adopting these practices can build a resilient workforce capable of meeting the demands of digital transformation.

A significant aspect of the study is its exploration of digital inequality—the potential for automation to displace routine jobs while benefiting highly skilled workers. Walkowiak argues that neurodiverse inclusion provides a solution to this disparity. By designing roles that leverage the unique strengths of neurodiverse individuals, organizations can address workforce displacement and ensure equitable access to the opportunities created by digital transformation.

The paper concludes by framing digital transformation as a pivotal opportunity for advancing neurodiversity in the workplace. By integrating autistic workers through targeted practices, businesses not only close skill gaps but also foster innovation and productivity. Walkowiak's research underscores the strategic value of neurodiversity in creating inclusive, efficient, and forward-thinking work environments in the digital era.

The RAND Corporation [40] explores the potential of neurodiversity to address the persistent talent shortage in national security. The paper emphasizes the distinct cognitive strengths of neurodivergent individuals—such as heightened focus, pattern recognition, and analytical thinking—as crucial for addressing complex challenges in intelligence, cybersecurity, and data analysis. Despite the demand for these skills, traditional hiring and workplace practices have largely excluded neurodiverse candidates, leaving a valuable talent pool untapped.

The study highlights examples of how neurodivergent traits align with national security roles. Tasks requiring sustained attention, anomaly detection, and long-term analysis benefit from the abilities of individuals with Autism Spectrum Disorder (ASD), ADHD, and dyslexia. Successful models, such as Israel's Unit 9900, demonstrate the effectiveness of neurodiverse teams in roles like satellite image analysis and cybersecurity threat monitoring. These examples showcase how inclusive hiring practices can strengthen operational capabilities.

Barriers to neurodiverse inclusion in national security are also examined. Misconceptions equating neurodivergence with disability often discourage candidates from disclosing their conditions or seeking accommodations. Additionally, rigid hiring standards, complex security clearance procedures, and workplace cultures that prioritize neurotypical traits over technical skills exacerbate exclusion. RAND advocates for revising these practices, recommending that job descriptions focus on essential technical competencies rather than social skills. Accommodations, such as sensory-friendly workplaces and alternative communication methods, are proposed to create supportive environments for neurodiverse employees.

To further support neurodivergent talent, RAND emphasizes the importance of continuous professional development and clear communication about workplace adjustments. These measures, the report argues, not only benefit neurodivergent staff but also improve overall workplace functionality. The paper concludes by asserting that integrating neurodiversity into national security is not merely a solution to skill shortages but a strategic advantage. By fostering diverse cognitive approaches, national security agencies can enhance their resilience, adaptability, and ability to counter emerging threats in an increasingly complex global landscape.

The role of neurodiversity in advancing government digital transformation has been explored by Hiren Shukla [41], highlighting its potential to enhance inclusivity and innovation. The concept of neurodiversity, which includes conditions such as autism, ADHD, and dyslexia, reflects the natural diversity in human cognition and is increasingly recognized for its workplace benefits. For government organizations, a neurodiverse workforce introduces varied thinking styles that contribute to the design of accessible tools and technologies, ultimately fostering public trust and credibility.

It has been emphasized that a workforce reflecting the diversity of the population strengthens the government's responsiveness and fairness. Studies indicate, however, that neurodivergent individuals face disproportionately high unemployment rates, with only 22% of autistic adults in the UK being employed. A similar trend exists in the U.S., demonstrating the urgent need for inclusive practices to support neurodivergent talent.

Research suggests that incorporating neurodiversity benefits digital transformation efforts by introducing fresh perspectives and cognitive approaches. For instance, autistic individuals are noted for their exceptional focus and pattern-recognition abilities, which are invaluable in data analysis and programming. A neurodiverse team is also found to be more resilient and adaptable, qualities critical in the rapidly evolving digital landscape. An example of successful integration is Israel's Milky Way program, which recruits neurodivergent individuals for cybersecurity roles, providing them with mentorship and accommodations to excel. This program has been recognized as a model for leveraging neurodiversity in government roles.

The integration of neurodiversity into digital transformation necessitates workplace adjustments, including flexible schedules, quiet workspaces, and neurodiversity training. Such initiatives not only represent a moral imperative but also serve as a strategic advantage. Governments that embrace neurodiversity can drive innovation and efficiency while fostering public trust, ensuring that digital transformation is inclusive and representative of the populations they serve.

Kashyap Kompella [42] explored the potential of artificial intelligence (AI) to address barriers faced by neurodivergent individuals in the workplace, emphasizing its role in fostering inclusivity and improving employment outcomes. Neurodiversity, which includes conditions such as autism, ADHD, and dyslexia, is increasingly recognized as a valuable asset in the workplace due to the unique perspectives and skills it brings. However, traditional work environments often pose challenges for neurodivergent individuals, necessitating innovative solutions.

The study highlighted the use of AI in enhancing diagnostics and education for neurodivergent individuals. Machine learning technologies have shown promise in analysing brain scans, speech patterns, and behaviours to diagnose conditions like autism with greater accuracy. In educational contexts, AI-enabled tools have allowed educators to customize teaching methods, thereby creating more inclusive and effective learning environments.

In workplace settings, Kompella [42] identified how AI can transform recruitment processes to better support neurodivergent candidates. Traditional hiring practices, often reliant on subjective assessments such as interpreting body language or maintaining eye contact, tend to disadvantage neurodivergent individuals. AI-based assessments, focusing on practical skills rather than interpersonal traits, provide a more equitable evaluation of candidates, helping organizations uncover hidden talent and foster diversity.

AI tools were also found to assist neurodivergent employees in adapting to workplace demands. Digital assistants can help manage schedules, set reminders, summarize meetings, and convert speech to text, enabling individuals to tailor tasks to their cognitive strengths. Moreover, remote and hybrid work models supported by AI provide sensory-friendly environments, addressing specific challenges neurodivergent employees might face in traditional office settings.

Kompella [42] underscored the ethical considerations in deploying AI for neurodiversity support. Ensuring that neurodivergent voices are included in the design and implementation of AI systems is critical to avoiding unintended biases and addressing unique challenges. Explicit inclusion of neurodiversity in AI ethics guidelines was advocated as a means to ensure fairness and effectiveness in technology deployment.

The study concluded that embracing AI as a tool for supporting neurodiversity is both a social responsibility and a strategic advantage. By leveraging AI to create inclusive and supportive workplaces, organizations can enable neurodivergent individuals to thrive, enhancing diversity, innovation, and productivity in the process.

Turel [43] et al. examined the role of neurodiversity among information systems (IS) users, particularly within leisure contexts, emphasizing the need for inclusivity in IS research and design. The study highlights that existing IS research predominantly focuses on neurotypical users in professional settings, often neglecting the varied ways neurodiverse individuals interact with systems in leisure domains, such as social media, e-commerce, and online reviews. With the increasing prevalence of neurodiverse conditions like ADHD, anxiety, and depression, the authors argue that overlooking these user groups risks marginalizing their experiences and creating systems that fail to address their unique needs. The study introduces a theoretical model to examine how neurodiversity influences decision-making in IS environments. The model identifies that neurodiverse users often engage in unique cognitive-emotional processes, shaping their reactions to perceived rewards, risks, and social interactions. Through three empirical studies, Turel and Bechara integrated aspects of neurodiversity into IS frameworks and identified notable behavioural differences. For instance, users with high anxiety were found to overestimate online privacy risks, while individuals with depression displayed lower levels of engagement and satisfaction with IS platforms. These findings underscore the necessity of accounting for neurodiverse traits to enhance the inclusiveness and accuracy of IS models.

The authors stress that designing IS systems with neurodiversity in mind can foster inclusivity and improve user experiences. They advocate for future IS research to systematically incorporate neurodiversity into its scope, ensuring that diverse cognitive and emotional processing styles are represented. By doing so, developers and researchers can enhance satisfaction, accessibility, and effectiveness, creating systems better suited to the needs of all users, including those with neurodiverse conditions.

Shrestha [44] et al. investigated the challenges faced by neurodiverse students with learning disabilities (LD) in online education, particularly concerning security, accessibility, and institutional support. The transition to online learning, accelerated by the COVID-19 pandemic, revealed significant gaps in preparedness and support for students with disabilities. Based on the experiences of 62 neurodiverse students, the study identified key issues related to privacy risks, cybersecurity awareness, and insufficient accommodations from educational institutions.

A prominent finding was the lack of adequate security measures and guidance provided to neurodiverse students. Participants reported struggles with navigating online privacy risks, such as phishing and other cyber threats, due to limited familiarity with cybersecurity practices. The absence of structured training and awareness programs heightened their vulnerability. Moreover, students expressed dissatisfaction with the support systems provided by their institutions, citing inadequate communication and a lack of proactive measures to address their needs during the shift to online education.

The study also highlighted accessibility challenges. Many students with LD noted that their institutions had not implemented the accommodations required for effective online learning. They reported feelings of being unprepared and unsupported, which negatively impacted their engagement and overall educational experience. These findings emphasize the need for tailored interventions to address both the security and accessibility needs of neurodiverse students.

Shrestha [44] et al. recommended systematic changes to improve the online learning environment for neurodiverse students. Key suggestions include thorough planning for online course transitions, the integration of cybersecurity training into onboarding processes, and the establishment of dedicated support systems. By implementing these measures, institutions can create safer and more inclusive online learning environments, ensuring that neurodiverse students feel both supported and secure. These adjustments are crucial for enhancing the educational experiences and outcomes of students with disabilities in a rapidly digitalizing academic landscape.

The reviewed studies collectively highlight the transformative potential of neurodiversity in enhancing innovation, inclusivity, and efficiency across diverse domains such as cybersecurity, digital transformation, national security, education, and information systems. By addressing the unique cognitive strengths of neurodiverse individuals, including heightened attention to detail and problem-solving abilities, these works underscore the strategic advantage of embracing neurodiversity to address skill gaps. The frameworks and methodologies proposed in these studies provide actionable insights into creating inclusive environments, from gamified training and AI-assisted recruitment to tailored workplace accommodations and accessible online learning systems.

A recurring theme across the literature is the necessity of rethinking traditional practices—be it recruitment, system design, or educational delivery—to better align with the diverse needs of neurodivergent populations. Empirical evidence demonstrates that leveraging neurodiverse strengths not only fosters innovation but also improves user satisfaction, operational performance, and inclusivity in both public and private sectors. Programs such as Israel’s Milky Way initiative and industry-driven efforts like SecureLD and the DXC Dandelion Program illustrate the tangible benefits of integrating neurodiversity into workforce and technology strategies.

Furthermore, these studies emphasize the ethical imperatives and practical benefits of inclusive practices, advocating for systemic changes that include tailored accommodations, cybersecurity awareness, and AI ethics that prioritize neurodiversity. By leveraging interdisciplinary approaches that integrate neuroscience, computer science, and organizational behaviour, this body of research lays a robust foundation for advancing neurodiversity initiatives. Collectively, these works contribute to a paradigm shift, positioning neurodiversity not just as a moral imperative but as a driver of innovation, equity, and competitive advantage in the digital and security-focused age.

Chapter 4. Dataset and Features

4.1 Dataset

In this research, a comprehensive framework is presented for the detection of ADHD using convolutional neural networks (CNNs) combined with advanced preprocessing and explainable artificial intelligence (XAI) techniques. The primary objective is to classify subjects into ADHD and control groups based on features extracted from multiple datasets.

The data utilized in this study were sourced from two widely recognized open-source EEG datasets, which were combined to form a new hybrid dataset [45]. This hybridization is a one of the contributions of this study, enhancing the accuracy of ADHD detection by leveraging the complementary strengths of both datasets and highlighting the relevance of specific features. The datasets underwent dimensional modifications to ensure proper alignment and consistency, facilitating a more robust analysis.

It is important to note that, to the best of our knowledge, no standardized benchmarking dataset currently exists for ADHD detection using EEG data. This study, therefore, represents a pioneering effort in the field, leveraging a hybrid dataset created from two distinct open-source datasets. The integration of these datasets enhances the model’s robustness by providing a diverse representation of ADHD-related neural activity. The absence of a benchmarking dataset highlights the innovative nature of this research, as it paves the way for future studies to adopt similar methodologies. While the lack of a benchmark poses challenges in comparative evaluations, it underscores the significance of the explainability techniques employed in this study, which aim to bridge the gap between data-driven models and their interpretability.

4.1.1 Data Gathering

The first dataset [14] consisted of EEG recordings from 40 adult participants, equally divided between 20 with ADHD and 20 healthy controls, each group comprising 10 males and 10 females. The data were collected while participants performed the Embedded Figures Task (EFT), a cognitive test designed to assess an individual's ability to locate simple geometric shapes hidden within complex, overlapping patterns. The EFT is widely used to evaluate selective attention, visual search skills, and cognitive flexibility, which are often impaired in individuals with ADHD. ADHD participants were diagnosed according to DSM-5 criteria by a certified clinician and had their symptoms evaluated using the Adult ADHD Self-Report Scale (ASRS-v1.1). Those undergoing stimulant treatment were instructed to discontinue medication two days before the experiment, following a physician-supervised protocol, with the option to resume afterward. Participants were excluded if they had psychosis, bipolar disorder, substance use disorder, or neurological conditions. Conversely, the control group participants were required to have no psychiatric or neurological conditions and to be free from psychoactive medications. All participants provided written informed consent, and the study was conducted in accordance with ethical guidelines, receiving approval from the Institutional Review Board of Partners HealthCare System, Massachusetts General Hospital.

The second dataset [15] involved EEG recordings from 121 children aged 7 to 12 years, including 61 diagnosed with ADHD and 60 healthy controls, encompassing both boys and girls. The ADHD group was diagnosed by a psychiatrist following DSM-V criteria and had been treated with Ritalin for a maximum of six months. The control group had no psychiatric or neurological conditions, epilepsy, or history of high-risk behaviors.

EEG signals were captured using a 19-electrode montage based on the 10–20 system at a sampling frequency of 128 Hz, with earlobe electrodes serving as references. This dataset was refined for use in an EEG competition hosted by the National Brain Mapping Laboratory (NBML), which included visual attention tasks involving cartoon characters. Children were asked to count the characters or perform a calculation based on the visual stimuli, with the duration of EEG recording dependent on their response speed. The modified version of this dataset, resampled to 512 Hz, was used for analysis. Epochs lasting 30 seconds were extracted for training and validation, with separate sets for unbiased conclusions.

4.1.2. Data Preprocessing

The EEGLAB [46] toolbox, a powerful open-source MATLAB [47] toolbox specifically designed for processing EEG data, was utilized to preprocess the datasets. EEGLAB [46] was employed to visualize and analyze the placement of electrodes on the human head according to the international 10-20 system, ensuring consistent electrode positioning across both datasets. The 10-20 system is a globally recognized standard for electrode placement on the scalp, designed to ensure consistent and reproducible EEG recordings. It divides the head into proportional distances based on anatomical landmarks, with electrodes spaced either 10% or 20% apart. This system labels electrodes by brain regions (e.g., F for frontal, T for temporal) and assigns odd numbers to the left hemisphere and even numbers to the right. By standardizing placement, the 10-20 system facilitates reliable comparisons across datasets and studies, making it a cornerstone in EEG research and clinical applications.

The toolbox facilitated an understanding of the dimensional structures of the datasets, enabling the identification and correction of any data inconsistencies. Furthermore, EEGLAB was used to convert the raw EEG data into a standardized format, such as CSV, allowing for seamless integration and combination of the two datasets. This conversion ensured proper alignment of all EEG signals, maintaining data integrity while making them readily accessible for subsequent machine learning analysis. By leveraging the advanced functionalities of EEGLAB, data quality was enhanced and compatibility across datasets was ensured, ultimately improving the robustness of the ADHD detection model.

4.1.3 Data Integration and Shuffling

After adjusting the dimensions, the datasets were concatenated into a single DataFrame, resulting in a unified dataset encompassing all samples. To prevent biases during model training and evaluation, the combined dataset was shuffled randomly.

Once the datasets were dimensionally adjusted, they were combined into a single unified DataFrame. This hybrid dataset integrated EEG recordings from adults and children, providing a comprehensive and diverse representation of ADHD-related neural patterns. To eliminate biases arising from data order and ensure randomness in the learning process, the combined dataset was shuffled prior to partitioning.

The unified dataset was split into training and testing sets following an 80-20 ratio. The training set, encompassing 80% of the data, was used to develop the CNN model by exposing it to a variety of patterns and relationships within the EEG data. The remaining 20% of the data constituted the testing set, reserved exclusively for final evaluation to ensure that the reported performance metrics were based on entirely unseen data. This separation between training and testing sets was critical for assessing the model's generalization capabilities, simulating real-world scenarios where the model would encounter new, previously unobserved data.

During training, an additional validation subset was extracted from the training data. This subset allowed for iterative refinement of the model by monitoring its performance and tuning hyperparameters, such as learning rates and regularization terms, while safeguarding against overfitting. The use of this three-way split—training, validation, and testing—provided a systematic framework to evaluate and improve the model at each stage of development.

Moreover, to address the inherent class imbalance in the ADHD and control groups, the Synthetic Minority Over-sampling Technique (SMOTE) was employed on the training data. By generating synthetic samples for the underrepresented class, SMOTE ensured that the model received balanced exposure to both ADHD and control data during training. This step was crucial for enhancing the model's sensitivity to ADHD-related patterns while preventing it from disproportionately favoring the majority class. This step was crucial to ensure that the order of the data did not influence the model's learning process.

To ensure the reliability and validity of the combined dataset, several data quality checks and preprocessing steps were performed. These steps included handling missing data, filtering out noise from EEG signals, and performing artifact correction to remove unwanted interference from non-brain activity, such as eye blinks or muscle movements. Band-pass filters were applied to retain frequencies of interest (typically 1-45 Hz) relevant to ADHD detection while discarding irrelevant frequencies. This filtering process helped enhance signal quality, ensuring that the data fed into the model was both clean and informative. Additionally, epoching techniques were used to segment continuous EEG recordings into smaller, time-locked intervals associated with specific events or tasks, further refining the dataset for model training.

4.2 Feature Extraction and Splitting

The independent variables (features X) and dependent variables (labels y) were extracted from the unified dataset. Subsequently, the data was split into training and testing sets using an 80-20 split ratio, ensuring a robust evaluation of the model's performance. Standardization of the features was performed to normalize the data, which was a critical step for optimizing the performance of the CNN. The training data was fitted and transformed using a StandardScaler, and the same transformation was applied to the test data. The unified dataset underwent further refinement to extract independent variables (features) and dependent variables (labels) required for model training. Features representing EEG signal characteristics and corresponding class labels were prepared for analysis. Before feeding this data into the CNN, it was standardized using a StandardScaler to normalize feature distributions. This ensured that variables across different scales contributed equally to the learning process, avoiding biases toward features with larger numerical ranges.

The training data was fitted and transformed during standardization, and the same transformation was applied to the testing data to maintain consistency across all phases. This preprocessing step ensured that the model learned effectively and could apply its knowledge to new, unseen samples without discrepancies caused by scale variations.

The testing set was completely withheld during model development to simulate real-world conditions. This unseen data served as the definitive benchmark for evaluating the model's generalization ability. Evaluation metrics such as accuracy, precision, recall, and F1-score were computed on the testing set to provide a comprehensive assessment of the model's classification capabilities. The exclusive use of the testing set for final evaluation eliminated any potential biases and validated the robustness of the proposed methodology.

Additionally, the segmented training data exposed the model to diverse ADHD-related patterns during the learning phase, enabling it to capture both broad trends and subtle nuances in EEG signals. This approach, combined with the incorporation of SMOTE and rigorous validation, ensured a balanced, accurate, and interpretable model for ADHD detection. To address class imbalances, the Synthetic Minority Over-sampling Technique (SMOTE) [48] [49] was applied to the training data, where applicable, to create a more balanced dataset for training. SMOTE generated synthetic samples for the underrepresented class, ensuring that the model received a balanced exposure to both ADHD and control data during training. This approach was crucial for improving the model’s sensitivity to ADHD-related patterns and preventing it from disproportionately favoring the majority class. By incorporating SMOTE, the training process was optimized for robust performance, addressing the challenges posed by imbalanced datasets while maintaining the integrity of the original data.

Chapter 5. Methodology

5.1 Model Architecture

The core of the methodology is the design and implementation of a 1D Convolutional Neural Network (CNN). The model architecture comprises an initial 1D convolutional layer with 32 filters and a kernel size of 3, followed by a max pooling layer with a pool size of 2. This structure is repeated with 64 filters in the second convolutional layer. The network then includes a flattening layer to convert the 2D matrix into a vector, which is passed through a dense layer with 64 neurons and a ReLU activation function. To mitigate overfitting, a dropout layer with a 50 percent dropout rate is added. The output layer consists of a single neuron with a sigmoid activation function, making it suitable for binary classification tasks. This architecture is chosen for its ability to capture temporal dependencies in the EEG data effectively.

5.2 Model Training and Evaluation

The model is compiled using the Adam optimizer and a custom focal loss function to address class imbalance, with accuracy as the evaluation metric. The focal loss function addresses class imbalance by assigning higher weights to misclassified samples, thus improving the model’s sensitivity to minority classes. Training is conducted over 30 epochs with a batch size of 32 and a validation split of 20 percent to monitor the model’s performance on unseen data. Early stopping and learning rate reduction callbacks are employed to prevent overfitting and optimize training. Post-training, the model’s accuracy on the test set is evaluated to ensure it generalizes well beyond the training data. These steps are essential to ensure that the model does not just memorize the training data but can also perform well on new, unseen data.

5.3 Prediction and Performance Metrics

Predictions are made on the test set, and probabilities are converted into binary class labels based on a threshold of 0.6. A comprehensive classification report (See Table 1), including precision, recall, and F1-score, is generated to provide detailed insights into the model's performance. Additionally, confusion matrices, ROC curves, and precision-recall curves are plotted to visually assess the model's classification capabilities. These metrics provide a clear picture of how well the model is distinguishing between ADHD and control subjects.

5.4 Explainability with LIME

Both LIME and SHAP were employed to enhance interpretability. LIME provided localized insights into individual predictions, while SHAP offered a global perspective on feature importance. Their combined use validated the model's reliability and transparency, ensuring clinical applicability.

A significant contribution of this study lies in its integration of explainability techniques, which provide transparency and insight into the CNN model's decision-making process. While LIME and SHAP have proven effective for interpreting model predictions, it is essential to note that explainability does not imply absolute accuracy or a complete understanding of the model's inner workings. These methods offer approximations of feature importance, helping identify the EEG channels and patterns most relevant to ADHD detection.

Explainability does not imply absolute clarity or completeness in understanding the model's operations. Machine learning models, especially neural networks, often function as "black boxes," with certain internal dynamics beyond human interpretation. Techniques like LIME offer insights by approximating the model's behavior for specific instances, but they cannot fully demystify every decision the model makes. For example, while LIME highlights influential EEG features, it may not accurately capture the interdependencies between these features or the contextual nuances that influence the model's broader predictions. This limitation is inherent to explainability frameworks, reflecting the balance between transparency and the innate complexity of machine learning algorithms. However, certain aspects of the machine learning model's behavior may remain unexplained due to the inherent complexity of neural networks. Despite this limitation, the use of explainability techniques ensures that the model's decisions are not only interpretable but also clinically meaningful, highlighting the critical features associated with ADHD while supporting trust in its predictions.

To enhance the interpretability of the model's predictions, we employ the LIME (Local Interpretable Model-agnostic Explanations) [12] technique. The LIME explainer is instantiated using the training data, feature names, and class names. Multiple instances from the test set are selected, and LIME generates explanations highlighting the most influential features contributing to the model's predictions. The aggregated feature importance is calculated and saved as an HTML file for further analysis. This step ensures that the model's decisions are transparent and interpretable. By understanding which features influence the model's predictions, we can gain insights into the underlying patterns in the EEG data that are associated with ADHD.

5.5 Explainability with SHAP

To enhance the interpretability of the model's predictions, we also employ the SHAP (SHAPley Additive explanations) technique. The SHAP explainer is instantiated using the trained model, providing a unified approach to measure feature importance based on cooperative game theory. SHAP values are computed for multiple instances from the test set, which represent the contribution of each feature to the prediction. These SHAP values are then aggregated to visualize [50] feature importance across the entire dataset. This step ensures that the model's decisions are not only transparent but also consistent and fairer across most of the predictions. By understanding which features have the greatest impact on the model's predictions, we can gain valuable insights into the EEG patterns associated with ADHD.

While SHAP provides a consistent global perspective on feature importance, it is essential to acknowledge that explainability techniques are inherently approximative. They rely on simplified representations of complex model behaviors, meaning certain interactions or subtleties within the model remain unaddressed. For instance, SHAP values distribute feature contributions based on cooperative game theory but may oversimplify the influence of non-linear interactions among features. This underscores the practical reality that explainable AI methods aim to enhance interpretability rather than deliver an exhaustive understanding of model operations. Despite these limitations, integrating SHAP into the diagnostic framework ensures clinicians and researchers can trust the model's outputs, even if some decisions remain partially opaque.

To enhance the interpretability of the model's predictions, we employ both the LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHAPley Additive explanations) techniques [12]. LIME provides local explanations by approximating the model for each specific instance, identifying the most influential features contributing to individual predictions. On the other hand, SHAP offers a consistent, global measure of feature importance, grounded in cooperative game theory, which ensures fairness by distributing feature importance proportionally by distributing the contribution of each feature proportionally based on its impact across all possible combinations of features. This method guarantees that each feature's importance is evaluated fairly, preventing any bias toward particular features or instances, thus providing a balanced and more equitable interpretation of the model's behavior across the entire dataset.

For our analysis, both LIME and SHAP explainers are initialized using the training data, feature names, and class labels. We selected multiple instances from the test set to generate explanations that highlight the key features driving the model's decisions. The aggregated feature importance from both LIME and SHAP is visualized [50] and saved for further analysis. This dual approach allows us to cross-verify the insights provided by each method, ensuring a robust interpretation of the model's behavior.

In the following chapter, we compare the results obtained from LIME and SHAP, analyzing any consistencies or discrepancies in the identified feature importance. This comparison provides a deeper understanding of the model's decision-making process and enhances the reliability of our interpretations of the EEG patterns associated with ADHD.

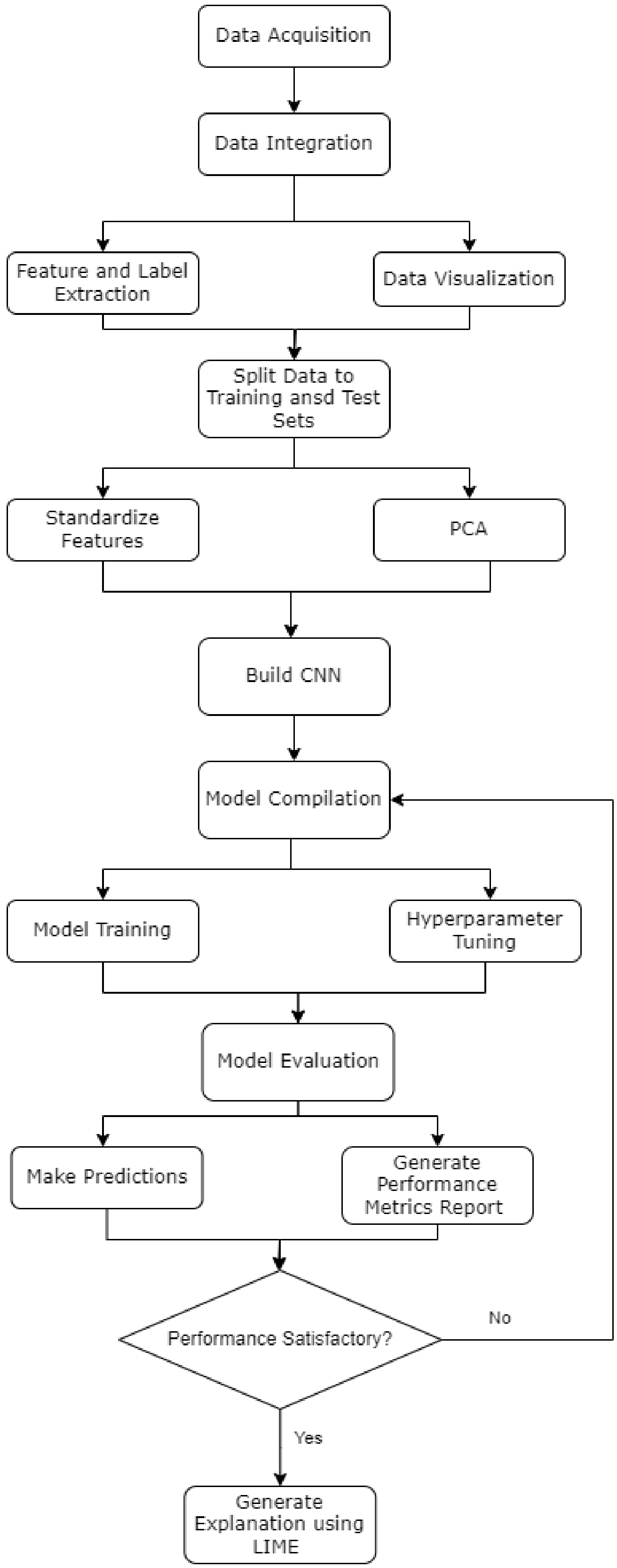


Figure 1: Flowchart illustrating the key steps in the ADHD detection methodology.

Chapter 6. Experimental Results

The results of the ADHD detection methodology, as illustrated by the visual outputs, underscore the exceptional performance of the convolutional neural network (CNN) model trained on the EEG data. The training process spanned 30 epochs, during which the model demonstrated consistent improvement, marked by a continuous decrease in loss and an increase in accuracy with each epoch. By the end of the training phase, the model reached an accuracy of 98.91 percent, indicating both strong learning capability and effective generalization to unseen data.

Table 1: Classification Report generated using the proposed model

Metric	**Value**
Test Accuracy	0.9891
Precision	0.64
Recall	0.88
F1-score	0.74

Upon evaluating the model on the test set, it achieved an impressive test accuracy of 98.91 percent. This high level of accuracy underscores the model's robustness and ability to generalize effectively beyond the training data, reinforcing its potential for practical application in ADHD detection. The classification report provides a detailed breakdown of the model's performance: for the ADHD group, the model demonstrated a precision of 0.64, indicating that 64% of the subjects predicted as having ADHD were correctly identified. The recall for the ADHD group was 0.88, highlighting the model's effectiveness in identifying a substantial proportion of actual ADHD cases. The resulting F1-score of 0.74 balances these metrics, suggesting a moderate trade-off between precision and recall.

The relatively high recall (0.88) is particularly crucial in a medical diagnostic context, where it is essential to identify as many true cases of ADHD as possible, minimizing the risk of missing any actual instances. Despite the relatively lower precision

(0.64), the model's high accuracy (98.91%) and recall (0.88) demonstrate its robustness and practical utility. In the context of medical diagnostics, where missing a true case (false negatives) can have severe consequences, prioritizing high recall is essential. The high recall ensures that most ADHD cases are identified, even if it means accepting a slightly higher rate of false positives. While this trade-off lowers precision, it aligns with the primary goal of medical diagnostics: minimizing the risk of undiagnosed cases. Additionally, the high overall accuracy indicates that the model performs well across the entire dataset, correctly identifying both ADHD and control cases. This balance of metrics underscores the model's suitability as a screening tool in clinical practice, where additional follow-up evaluations can address false positives, but missed diagnoses could have significant negative outcomes.

Overall, the classification report reflects a strong performance across various metrics, and the near-perfect test accuracy highlights the model’s capability to distinguish between ADHD and control subjects effectively. The findings suggest that while the model is highly sensitive in detecting ADHD cases, optimizing for precision remains an important focus to further refine its diagnostic utility.

Furthermore, while some methods like CNN with classical classifiers of raw EEG data [28] achieved high accuracy (95.83%), they still fall short of the performance of our proposed methodology. The superior accuracy of the proposed method can be attributed to its effective integration of multiple datasets, sophisticated feature extraction techniques, and robust neural network architecture, which collectively provide a more comprehensive understanding and detection of ADHD. These results demonstrate that the proposed approach offers a more powerful and reliable solution compared to existing methodologies, supporting its potential for clinical application in ADHD diagnosis.

6.1 Receiver Operating Characteristic (ROC) Curve

The Receiver Operating Characteristic (ROC) curve, depicted in Figure 2, provides a graphical representation of the model's performance by plotting the true positive rate against the false positive rate. The area under the curve (AUC) is 0.97, reflecting the model's strong ability to distinguish effectively between the ADHD and control groups.

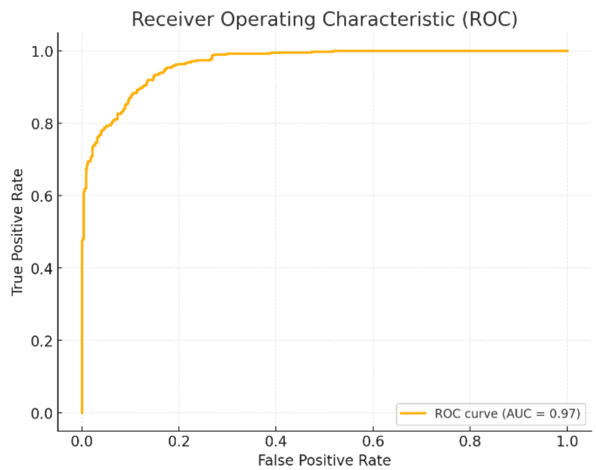


Figure 2: Receiver Operating Characteristic Curve

With an accuracy of 98.91%, a precision of 0.64, and a recall of 0.88, the model demonstrates a balance between sensitivity and specificity. The F1-score of 0.74 further confirms the model's ability to achieve a reasonably high true positive rate while managing the trade-off with false positives. These results highlight the model's reliability and accuracy in ADHD detection, suggesting robust performance in practical diagnostic settings, with potential for further optimization.

6.2 Precision-Recall Curve

Figure 3 presents the precision-recall curve, which offers a detailed examination of the trade-off between precision and recall. The curve illustrates how precision varies with different levels of recall, highlighting the model's ability to maintain high recall levels while managing varying precision. Notably, precision is highest at lower recall values and decreases as recall increases, reflecting the model's performance dynamics across different decision thresholds. The curve shows a recall of 0.88 and a precision of 0.64, indicating the model's sensitivity and trade-off in practical settings. The curve also provides insight into the model's behavior under varying conditions, revealing the impact of different classification thresholds. By analyzing this curve, one can determine the optimal balance between precision and recall for specific application needs.

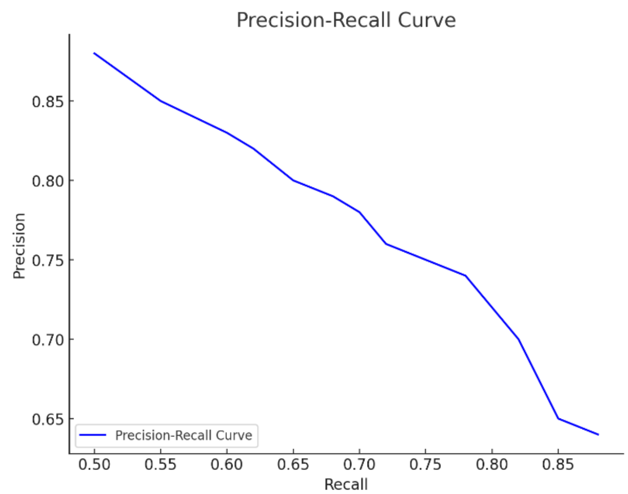


Figure 3: Precision-Recall Curve

6.3 Training and Validation Loss/Accuracy

The graph in Figure 4 shows the training and validation loss/accuracy over the course of

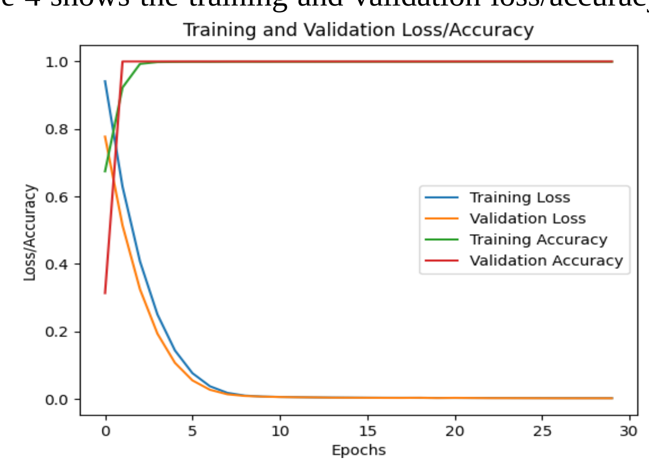


Figure 4: Training and Validation Loss-Accuracy

the training epochs. Both the training and validation losses decrease steadily, converging to low values, while the training and validation accuracy increase, approaching near-perfect levels. This trend indicates that the model is effectively learning from the training data and generalizing well to the validation data, demonstrating minimal overfitting and strong performance across both datasets.

6.4 Aggregated Feature Importance

The aggregated feature importance visualizations (Figures 5 and 6) offer a comprehensive overview of the contributions of different features to the model's predictions, as interpreted by the LIME and SHAP techniques. These graphs quantify the influence of individual features on the model's decision-making process. LIME (Local Interpretable Model-agnostic Explanations) provides a framework for explaining the model's behavior by identifying critical features influencing specific predictions. The aggregated importance from LIME is calculated by averaging the importance scores across multiple instances, thereby highlighting which features consistently play a significant role in distinguishing between ADHD and control subjects.

SHAP (SHAPley Additive explanations) provides a global perspective on feature

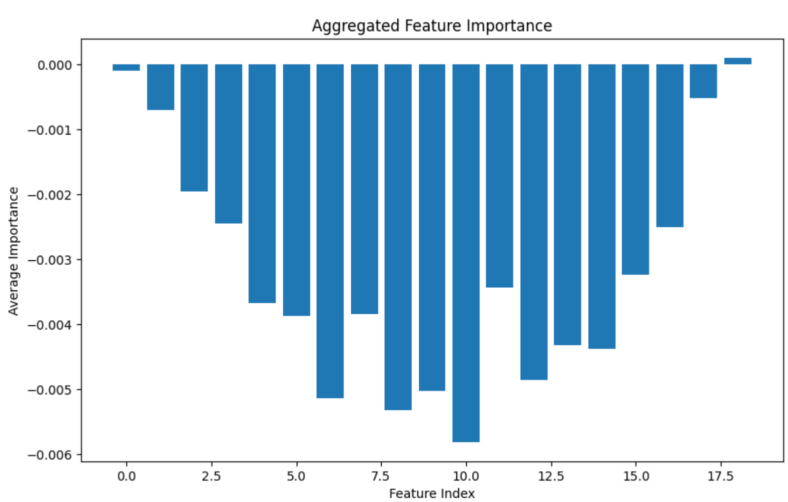


Figure 5: Aggregated Feature Importance using LIME

importance based on cooperative game theory. SHAP values ensure fairness by assigning each feature a contribution to the prediction that is consistent across all instances, based on

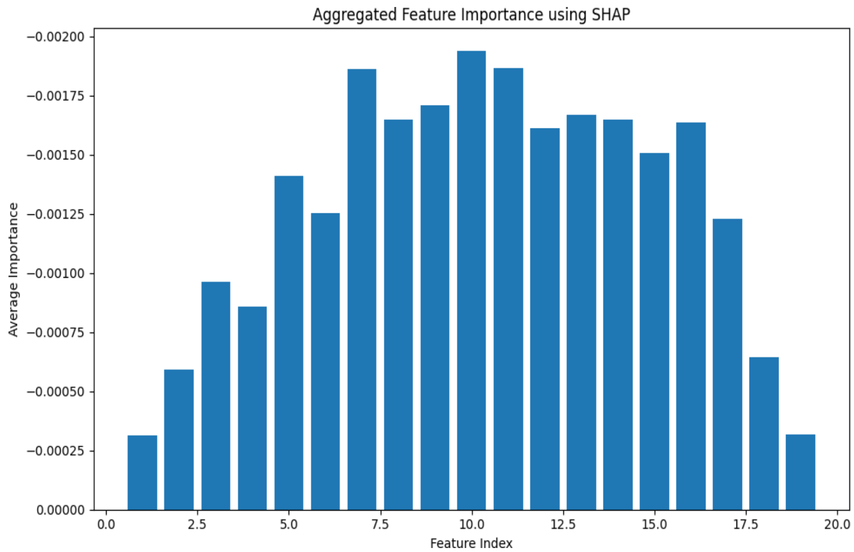


Figure 6: Aggregated Feature Importance using SHAP

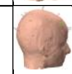
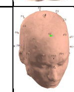
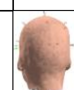
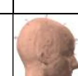
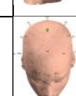
its impact when considered alongside all other features. This approach guarantees that the contribution of each feature is calculated in a balanced and unbiased manner, allowing for a comprehensive understanding of the model's behavior across the entire dataset.

By comparing the insights derived from SHAP and LIME, we can validate the significance of specific EEG features in differentiating ADHD from control subjects. The following table (Table 2) presents a detailed comparison of the importance scores assigned by both SHAP and LIME across various EEG channels. These scores indicate the relative contribution of each channel's signal to the model's predictions, with higher values signifying greater importance.

As observed in Table 2, the alignment between SHAP and LIME importance scores across multiple EEG channels confirms the robustness of the model's interpretability. Notably, both methods identify channels P3, Cz, and C4 as highly influential in classifying ADHD, with LIME assigning the highest importance to channel Cz, while SHAP highlights P3 and C4 as key contributors. This consistency between the two XAI methods suggests that these channels play a pivotal role in the model's decision-making process, providing meaningful insights into the differentiation of ADHD from control subjects.

Furthermore, the visual snapshots in the table (Table 2) offer additional evidence, displaying the EEG patterns corresponding to these key features. These visualizations help understand the underlying neural activities captured by different EEG channels by highlighting the spatial distribution and intensity of the neural signals across the scalp. By visualizing the activation patterns, it becomes easier to interpret which specific brain regions and corresponding channels are most involved in distinguishing between ADHD and control subjects, providing a more intuitive understanding of the model's decision-making process.

Table 2: Comparison of SHAP and LIME Importances with Snapshots

Channel Name	**Label**	**SHAP Importance**	**LIME Importance**	**Snapshot**
1	Fp1	0.000312	0.000383	
2	Fz	0.000592	0.000961	
3	Fp2	0.000963	0.001032	
4	F3	0.000858	0.001087	
5	F4	0.001411	0.001570	
6	C3	0.001251	0.001628	
7	C4	0.001863	0.001643	
8	P3	0.001649	0.002482	
9	P4	0.001707	0.002019	
10	O1	0.001937	0.001810	
11	O2	0.001864	0.001522	
12	F7	0.001612	0.001761	
13	F8	0.001668	0.001841	
14	T7	0.001649	0.001921	
15	T8	0.001508	0.001129	
16	P7	0.001634	0.000648	
17	P8	0.001228	0.000238	
18	Fz	0.00064616	0.000111	
19	Cz	0.000318	0.006513	

The snapshot of the head (in Table 2) is oriented to maintain the corresponding channel in focus. Figure 7, shows a zoomed-in view of EEG electrode placement on the human head, used for generating EEG signals with EEGLAB.

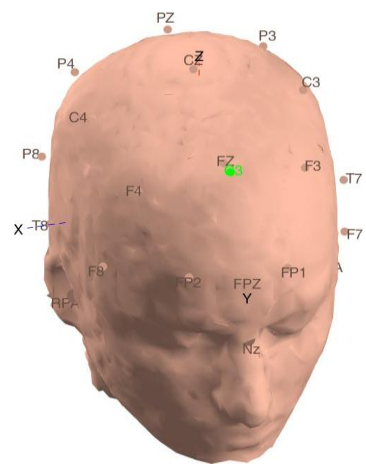


Figure 7: Sample Snapshot of EEG Electrode Placement on Human Head for Generating EEG Signals using EEGLAB from MATLAB toolbox

The consistent predictions across multiple runs of the model further validate its reliability and robustness. The explanations generated by LIME are saved in an HTML format, offering comprehensive insights into the model's reasoning process and enhancing transparency. This documentation ensures that stakeholders can trust the results and establishes a foundation for future research.

Chapter 7. Conclusion

This chapter presents a comprehensive summary of the contributions, findings, and implications of this thesis, which investigated the application of advanced machine learning techniques, particularly convolutional neural networks (CNNs), alongside explainable AI (XAI) methodologies for ADHD detection using EEG data. The study sought to address gaps in interpretability and diagnostic reliability, providing a pathway for integrating robust AI systems into clinical settings. Section 7.1 outlines the key contributions and innovations developed through the course of this research, while Section 7.2 discusses potential future directions, highlighting opportunities for expanding and refining the methodologies proposed in this thesis.

7.1 Contributions

The contributions of this thesis lie at the intersection of machine learning, neuroscience, and health informatics, aiming to advance the state of ADHD detection and its clinical utility. The predominant reliance on traditional classifiers and insufficient attention to explainability highlighted the need for a novel approach. To address these gaps, this research developed a comprehensive framework incorporating robust data preprocessing, CNN-based modeling, and a comparative analysis of two prominent XAI techniques: Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP). The methodology introduced in this thesis demonstrated superior performance compared to existing approaches by achieving high diagnostic accuracy and interpretability. CNN models were optimized to process EEG data effectively, leveraging both spatial and temporal patterns indicative of ADHD. The inclusion of SHAP analysis provided global consistency in understanding feature importance across the entire dataset, offering clinicians a clear and balanced view of the key EEG features contributing to model predictions. In contrast, LIME's localized interpretability enabled precise insights into model behavior for specific instances, complementing the broader perspective offered by SHAP. Together, these techniques established a robust framework for balancing performance with explainability. Explainability plays a dual role in enhancing machine learning models for ADHD detection. First, it bridges the gap between model complexity and clinical trust by providing interpretable insights into EEG patterns. Techniques like LIME and SHAP uncover relationships between input features and predictions, ensuring that clinicians can verify and understand the decision-making process. However, it is crucial to recognize that explainability does not equate to exhaustive clarity. Machine learning models, particularly convolutional neural networks, rely on abstract representations that may not align perfectly with human interpretability. While these tools reveal influential patterns, the underlying interactions within the model often remain hidden, reflecting the inherent trade-offs in explainable AI. This limitation underscores the importance of positioning explainability as a supportive tool rather than a definitive explanation of all model decisions.

This work advances the understanding of ADHD-related neural patterns, identifying critical EEG features that strongly influence diagnostic outcomes. By combining high-performing CNN models with interpretable XAI methods, this research contributes to the growing field of transparent AI in healthcare. It provides a scalable and reliable approach that can be adapted to other neurodevelopmental disorders, addressing a critical need for actionable AI tools in clinical environments.

7.2 Future works

While this thesis provides a significant contribution to ADHD detection, several promising avenues for future research are identified. These opportunities aim to build on the current work, exploring both methodological enhancements and broader applications.

1. **Integration of Advanced Neural Architectures**: Future research can explore incorporating advanced neural network architectures such as transformers or graph neural networks (GNNs). Transformers, with their ability to model long-range dependencies in sequential data, could enhance the detection of subtle ADHD patterns in EEG signals. Similarly, GNNs could be applied to EEG data represented as connectivity graphs, capturing complex relationships between brain regions to improve predictive performance.
2. **Expanding Data Diversity and Volume**: To generalize the proposed methodology across diverse populations, future studies should involve larger and more diverse datasets. Expanding the dataset could help uncover cross-demographic variations in ADHD-related neural activity. Additionally, utilizing pre-trained models on large-scale neurological datasets could improve transfer learning approaches, particularly in low-data scenarios.
3. **Refining Explainability Techniques**: While SHAP and LIME provided significant insights, further exploration of hybrid XAI techniques could enhance the balance between local and global interpretability. Integrating domain-specific priors or clinical constraints into explainability models could make these tools even more relevant for healthcare professionals.

4. **Real-World Deployment and Validation**: Future efforts should focus on transitioning the proposed framework to real-world clinical settings. This involves validating the model's performance with clinician feedback, ensuring usability, and addressing potential challenges in deploying AI systems in medical environments. Collaborations with healthcare providers will be essential to refine the framework for practical applications.
5. **Exploring Multimodal Data Integration**: Combining EEG data with other modalities, such as MRI or genetic data, could provide a more holistic understanding of ADHD. A multimodal approach could enhance model robustness and provide deeper insights into the biological underpinnings of the disorder.
6. **Ethical and Social Considerations**: As AI models gain prominence in healthcare, addressing ethical concerns around data privacy, bias, and fairness will be critical. Future research must address data privacy and fairness, ensuring that AI tools are equitable and transparent for clinical use.

7.3 Final Remarks

In conclusion, this thesis successfully demonstrates a novel and effective framework for ADHD detection using EEG data, integrating state-of-the-art CNN modeling with explainable AI techniques. The findings contribute significantly to the development of reliable and interpretable diagnostic tools, advancing both the technical and clinical aspects of health informatics. By bridging the gap between machine learning and neuroscience, this research fosters a paradigm shift toward more precise, transparent, and actionable AI-driven healthcare solutions, laying the groundwork for improved diagnostic accuracy and patient outcomes.

References

[1] J. E. Santiago, “No Title.” https://www.sans.org/blog/primer-on-neurodiversity-in-cybersecurity/.

[2] S. B. Melodie Wilson, “No Title,” [Online]. Available: https://netflixtechblog.medium.com/embracing-neurodiversity-within-information-security-ed53fe9518fa.

[3] S. P. Dame Vivian Hunt, Dennis Layton, “No Title,” *Why Divers. matters by McKinsey Co.*, [Online]. Available: https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/why-diversity-matters.

[4] “No Title,” [Online]. Available: https://neurodiversehackers.com/.

[5] F. L. Britton JW, Frey LC, Hopp JLet al., authors; St. Louis EK, “Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants,” [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK390354/.

[6] A. Saxena, “An Introduction to Convolutional Neural Networks,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 10, no. 12, pp. 943–947, 2022, doi: 10.22214/ijraset.2022.47789.

[7] B. TaghiBeyglou, A. Shahbazi, F. Bagheri, S. Akbarian, and M. Jahed, “Detection of ADHD cases using CNN and classical classifiers of raw EEG,” *Comput. Methods Programs Biomed. Updat.*, vol. 2, no. November, p. 100080, 2022, doi: 10.1016/j.cmpbup.2022.100080.

[8] 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.*, vol. 16, no. 6, 2019, doi: 10.1088/1741-2552/ab3a0a.

[9] 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.*, vol. 14, no. April, pp. 1–12, 2020, doi: 10.3389/fnins.2020.00251.

[10] D. Tanko and others, “EPSPatNet86: Eight-pointed star pattern learning network for detection ADHD disorder using EEG signals,” *Physiol. Meas.*, vol. 43, no. 3, 2022, doi: 10.1088/1361-6579/ac59dc.

[11] N. Ahire, R. N. Awale, and A. Wagh, “Comprehensive review of EEG data classification techniques for ADHD detection using machine learning and deep learning,” *Rom. J. Pediatr.*, vol. 72, no. 2, pp. 57–66, 2023, doi: 10.37897/RJP.2023.2.1.

[12] M. T. Ribeiro, S. Singh, and C. Guestrin, “‘Why Should I Trust You?’ Explaining the Predictions of Any Classifier,” *NAACL-HLT 2016 - 2016 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. Proc. Demonstr. Sess.*, pp. 97–101, 2016, doi: 10.18653/v1/n16-3020.

[13] S. M. Lundberg and S. I. Lee, “A unified approach to interpreting model

predictions," *Adv. Neural Inf. Process. Syst.*, vol. 2017-December, no. Section 2, pp. 4766–4775, 2017.

[14] A. Motie Nasrabadi, A. Allahverdy, M. Samavati, and M. R. Mohammadi, "EEG data for ADHD / Control children." IEEE Dataport, 2020, doi: 10.21227/rzfh-zn36.

[15] G. Sadeghi Bajestani, S. Abedian, F. Makhloughi, M. Raoufitabar, and H. Saeedi, "A Dataset of EEG Signals from Adults with ADHD and Healthy Controls: Resting State, Cognitive function, and Sound Listening Paradigm." Mendeley Data, 2023, doi: 10.17632/6k4g25fhzg.1.

[16] M. Treadwell, "Why Cybersecurity Benefits from Neurodiversity." https://www.beyondtrust.com/blog/entry/why-cybersecurity-benefits-from-neurodiversity.

[17] J. Asbell-Clarke, "How Neurodiversity Can Help Fill the Cybersecurity Workforce Shortage," [Online]. Available: https://www.darkreading.com/cybersecurity-operations/how-neurodiversity-can-help-cybersecurity-workforce-shortage.

[18] J. Asbell-Clarke, "How Neurodiversity Can Help Fill the Cybersecurity Workforce Shortage," [Online]. Available: https://careerhub.students.duke.edu/blog/2024/08/30/how-neurodiversity-can-help-fill-the-cybersecurity-workforce-shortage/.

[19] ISC2, "Cybersecurity Workforce Study," *Forrester Res.*, pp. 1–86, 2022, [Online]. Available: www.isc2.org/research.

[20] O. Pickup, "Why cybersecurity leaders are actively recruiting neurodiverse talent," [Online]. Available: https://www.worklife.news/technology/why-cybersecurity-leaders-are-actively-recruiting-neurodiverse-talent/.

[21] H. Foxcroft, "Spiky profiles - neurodiversity."

[22] P. Carpenter, "The Official Cybersecurity Summit," [Online]. Available: https://cybersecuritysummit.com/speaker/carpenter-perry/?utm_source=chatgpt.com.

[23] R. Page, "Talent overlooked: embracing neurodiversity in cybersecurity." https://www.csoonline.com/article/3616024/talent-overlooked-embracing-neurodiversity-in-cybersecurity.html?utm_source=chatgpt.com.

[24] P. Jarrett, "Neurodiversity and operating systems explained for people who get computers," [Online]. Available: https://www.linkedin.com/pulse/neurodiversity-operating-systems-explained-people-who-pete-jarrett/.

[25] S. Gauthier, "You're Not Disabled Enough," [Online]. Available: https://www.stem-sadie.com/post/you-re-not-disabled-enough.

[26] J. E. Robison, "Look Me in the Eye," [Online]. Available: https://www.goodreads.com/book/show/454856.Look_Me_in_the_Eye.

[27] M. Haynes, "Closing the Cybersecurity Gap: How Neurodiversity Can Be Your Secret Weapon." https://tofflerassociates.com/vanishing-point/closing-the-cybersecurity-gap-how-neurodiversity-can-be-your-secret-weapon/.

[28] B. TaghiBeyglou, A. Shahbazi, F. Bagheri, S. Akbarian, and M. Jahed, "Detection of ADHD cases using CNN and classical classifiers of raw EEG," *Comput. Methods Programs Biomed. Updat.*, vol. 2, no. November, p. 100080, 2022, doi: 10.1016/j.cmpbup.2022.100080.

[29] H. W. Loh *et al.*, "Deep neural network technique for automated detection of ADHD and CD using ECG signal," *Comput. Methods Programs Biomed.*, vol. 241, no. May, p. 107775, 2023, doi: 10.1016/j.cmpb.2023.107775.

[30] M. Duda, R. Ma, N. Haber, and D. P. Wall, "Use of machine learning for behavioral distinction of autism and ADHD," *Transl. Psychiatry*, vol. 6, no. 2, pp. 1–5, 2016, doi: 10.1038/tp.2015.221.

[31] S. K. Khare and U. R. Acharya, "An explainable and interpretable model for attention deficit hyperactivity disorder in children using EEG signals," *Comput. Biol. Med.*, vol. 155, no. October 2022, 2023, doi: 10.1016/j.compbiomed.2023.106676.

[32] I. Hussain *et al.*, "An Explainable EEG-Based Human Activity Recognition Model Using Machine-Learning Approach and LIME," *Sensors*, vol. 23, no. 17, 2023, doi: 10.3390/s23177452.

[33] D. Kuang and L. He, "Classification on ADHD with deep learning," in *Proc. - 2014 Int. Conf. Cloud Comput. Big Data, CCBD 2014*, 2014, pp. 27–32, doi: 10.1109/CCBD.2014.42.

[34] F. C. Morabito, C. Ieracitano, and N. Mammone, "An explainable Artificial Intelligence approach to study MCI to AD conversion via HD-EEG processing," *Clin. EEG Neurosci.*, vol. 54, no. 1, pp. 51–60, 2023, doi: 10.1177/15500594211063662.

[35] J. Scanlan *et al.*, "Neurodiverse Knowledge, Skills and Ability Assessment for Cyber Security," *ACIS 2020 Proc. - 31st Australas. Conf. Inf. Syst.*, 2020.

[36] "Neurodiversity in the technical security workplace."

[37] B. K. Wiederhold, "Diverse Minds, Secure Networks: Embracing Neurodiversity in Cybersecurity," *Cyberpsychology, Behav. Soc. Netw.*, vol. 27, no. 9, pp. 613–615, 2024, doi: 10.1089/cyber.2024.0413.

[38] K. C. Demek, B. Giunta, and R. Pinsker, "Bridging the gap in talent: A framework for interdisciplinary research on autism spectrum disorder persons in accounting and information systems," *Int. J. Account. Inf. Syst.*, vol. 55, no. May, 2024, doi: 10.1016/j.accinf.2024.100712.

[39] E. Walkowiak, "Neurodiversity of the workforce and digital transformation: The case of inclusion of autistic workers at the workplace," *Technol. Forecast. Soc. Change*, vol. 168, no. March 2020, p. 120739, 2021, doi: 10.1016/j.techfore.2021.120739.

[40] C. Weinbaum, O. Khan, T. D. Thomas, and B. D. Stein, *Neurodiversity and National Security: How to Tackle National Security Challenges with a Wider Range of Cognitive Talents*. 2023.

[41] H. Shukla, "Neurodiversity: An Important Contributor to the Government's Pivot

through Digital Transformation," *Pivoting Gov. through Digit. Transform.*, pp. 73–80, 2023, doi: 10.1201/9781003369783-5.

[42] A. I. F. O. R. Diagnosis, "How AI Can Help Companies Embrace Neurodiversity," no. august, pp. 39–41, 2023.

[43] O. Turel and A. Bechara, "Time to Talk About It: Neurodiversity of Information Systems Users in Leisure Settings," *J. Assoc. Inf. Syst.*, vol. 25, no. 5, pp. 1343–1371, 2024, doi: 10.17705/1jais.00877.

[44] S. Shrestha, D. Thomas, and S. Das, "SecureLD: Secure And Accessible Learning for Students with Disabilities," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 66, no. 1, pp. 465–469, 2022, doi: 10.1177/1071181322661157.

[45] A. B. Pedapaga, "ADHD Detection." [Online]. Available: https://github.com/akashpedapaga/ADHD-Detection.

[46] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics," *J. Neurosci. Methods*, vol. 13, pp. 9–21, 2004.

[47] "MATLAB," [Online]. Available: https://www.mathworks.com/products/matlab.html.

[48] C. K. A. Guillaume Lemaˆıtre, Fernando Nogueira, "Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning," [Online]. Available: https://jmlr.org/papers/volume18/16-365/16-365.pdf.

[49] S. Kiyohara, T. Miyata, and T. Mizoguchi, "Prediction of grain boundary structure and energy by machine learning," vol. 18, pp. 1–5, 2015, [Online]. Available: http://arxiv.org/abs/1512.03502.

[50] J. D. Hunter, : "a 2D G," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007.