Independent University **Bangladesh (IUB)**

IUB Academic Repository

Computer Science and Engineering | Undergraduate Thesis

2026-04

SHUCHOK: Early Identification of ADHD in Bangladeshi Children Using Machine Learning Models and a Questionnaire-Based App Interface

Panday, Sharbany

IUB

https://ar.iub.edu.bd/handle/11348/1199

Downloaded from IUB Academic Repository

SHUCHOK: Early Identification of ADHD in Bangladeshi Children Using Machine Learning Models and a Questionnaire-Based App Interface

April 2026

Prepared by:

Sharbany Panday (ID: 2221531)

Sourish Bairagi (ID: 2222817)

Sinthia Sayma Rahman (ID: 2221969)

Department of Computer Science and Engineering

Independent University, Bangladesh

Supervised by

Ashraful Islam, PhD

Assistant Professor

Department of Computer Science and Engineering

Independent University, Bangladesh

Co-supervised by

Farhana Sarker, PhD

Associate Professor

Department of Computer Science and Engineering

Independent University, Bangladesh

Thesis submitted as a requirement for the degree of Bachelor of Science (BSc) in Computer Science and Engineering (CSE) at Independent University, Bangladesh

Attestation

We are aware of the fact that plagiarism is strictly prohibited, and it conflicts with our university's rules and regulations. We guarantee the authenticity of our work. We have used some copyrighted material and models of others in our project, which have been properly cited following the international standards and proper guidelines.

Author Name: Sharbany Panday
Signature: _____

Author Name: Sourish Bairagi
Signature: _____

Author Name: Sinthia Sayma Rahman
Signature: _____

Evaluation Committee

Supervision Panel

.....	.....
Dr. sup	Dr. co sup
Supervisor	Co-supervisor

Examiners

.....	.....
Dr. internal (Internal)	Mr. internal 2 (Internal)

.....

Mr. external (External Examiner)

Office Use

.....	.....
Dr. AKM Mahbubur Rahaman	Dr. Saadia Binte Alam
Head of the Department	Program Coordinator

.....

Prof. Dr. M. Arshad Momen
Director, Office of the Graduate Studies & Research in Information and Resources.

.....

Library

Acknowledgment

We would like to extend our heartfelt thanks to our dear supervisor, Dr. Ashraful Islam and co-supervisor Dr. Farhana Sarker from Independent University, Bangladesh. Their continuous support, valuable suggestions, precious time and immense patience have been instrumental in the successful execution of our project. Their guidance and support helped us in making our work more enriching and we are indebted to them for their continued support in our educational career. Furthermore, we would like to express our sincere gratitude to the Center for Computational and Data Sciences (CCDS) for their invaluable assistance and support. The collaborative, helpful environment and resources at CCDS played a vital role in our project. Further, the support and guidance given to us has greatly contributed to the development of our academic and research careers. We sincerely appreciate the support and resources provided, which have greatly strengthened our research and broadened our understanding. We are also grateful to our parents for their unconditional support, encouragement and understanding. Their support and encouragement have been invaluable to us. Lastly, we recognize the utilization of artificial intelligence tools and academic search engines, such as ChatGPT (OpenAI), Claude (Anthropic) and Gemini (Google), in helping to support certain areas of this work. These tools were mostly used as help in coding, formatting and language polishing. We independently developed all of the fundamental elements of this report, such as the literature review, methodology, results, analysis, and discussion. We would also like to thank all those who helped us, directly or indirectly, in this project.

Abstract

The study focuses on attention deficit hyperactivity disorder that is underdiagnosed in Bangladeshi children, a critical public health concern. Due to the lack of awareness, high levels of social stigma and a severe shortage of available and culturally-appropriate screening methods, many children with signs of ADHD are undiagnosed, thus impeding their behavioral and academic progress. This research aims to bridge this gap by developing a culturally sensitive, economical and data-driven machine learning algorithm to detect signs of ADHD in the early stages, and to develop an app interface of ADHD screening (SHUCHOK) in the Bangladeshi setting.

Our method included a collection of a survey-based data set based on 110 participants between the ages of 4 and 17 with diagnosis of ADHD (n = 44) and without (n = 66). A DSM-5-based questionnaire translated into Bengali was used to capture behavioral symptoms, demographics, developmental history, and environmental factors. The four machine learning models were Random Forest, Support Vector Machine, Logistic Regression and Decision Tree which were trained and evaluated after subjecting the data to extensive data preprocessing and feature selection approaches, including Recursive Feature Elimination and Select Percentile.

The Random Forest model had the highest accuracy (91%), precision (0.90), recall (0.88), F1-Score (0.90) and Area Under the Curve (0.95) when compared to the others. In this sample, the most important predictive characteristics of ADHD are inattention, fidgeting, study focus and level of irritation of the child.

The main implication of this research is that it has great potential to enhance public awareness and early diagnosis of ADHD in low-resource countries such as Bangladesh. This study will significantly improve the accessibility of diagnoses and timeliness of interventions by offering a machine learning-based screening tool, which is proven to enhance the developmental outcome of the children affected.

Contents

1 Introduction **9**

1.1 ADHD Overview 9
1.2 Problem Statement 11
1.3 Objectives 12
1.4 Key Contribution of This Study 13
1.5 Research Questions 13
1.6 Limitations and Challenges 14

2 Literature Review **15**

2.1 Materials and Methods 15
2.2 Related Work 16
2.3 Literature Search Outcome 20
2.4 ML Model Popularity 22
2.5 Heatmap and Model Usage 23
2.6 Benchmarking Model Performance 24

3 Methodology **29**

3.1 System Architecture 29
3.2 Dataset Description 32
3.2.1 Survey Design 32
3.2.2 Ethical Approval 34
3.2.3 Data Collection 34
3.2.4 Tools and Technologies 39
3.3 Data Preprocessing and Feature Selection 39
3.4 Data Splitting 44
3.5 ML Model Training 46
3.6 Model Evaluation 53

4 App Interface **55**

4.1 SHUCHOK Overview 55
4.2 Interface Design 56

4.3 Prototype 58

5 Results and Analysis **61**

5.1 Performance Analysis of Random Forest Model 61
5.1.1 Error Analysis via Confusion Matrices 61
5.1.2 Statistical Evaluation of Performance Metrices 62
5.1.3 Probabilistic Interpretation (ROC/AUC) 63
5.2 Performance Analysis of Support Vector Machine (SVM) 64
5.2.1 Detailed Error Analysis (Confusion Matrices) 64
5.3 Quantitative Statistical Evaluation 64
5.3.1 ROC Curve and Discriminatory Analysis 66
5.4 Performance Analysis of Logistic Regression 67
5.4.1 Error Analysis and Classification Distribution 67
5.4.2 Statistical Evaluation of Quantitative Metrics 67
5.4.3 AUC and ROC Curve Interpretation 69
5.5 Performance Analysis of Decision Tree 70
5.5.1 Analysis of Classification Errors (Confusion Matrices) 70
5.5.2 Quantitative Metrics and Statistical Summary 70
5.5.3 ROC Curve and Discriminatory Analysis 72
5.6 Comparative Analysis of All Models 73
5.6.1 Final Cross-Validation Summary 73
5.6.2 Multi-Model ROC Curve Analysis 74
5.7 Final Model Selection and Justification 74
5.8 Feature Importance 76

6 Discussion **79**

6.1 Significance of Findings 79
6.2 Comparison with Existing Work 80
6.3 Significance of Interface 81
6.4 Limitations 82
6.5 Practical Implications in the Bangladeshi Settings 84

7 Conclusion **86**

8 Future Work **88**

Bibliography **96**

List of Figures

1.1 ADHD Brain Vs Normal Brain [Source] 11
1.2 Worldwide ADHD Overview [Source] 11

2.1 PRISMA flow diagram 17
2.2 Conceptual Heatmap 28

3.1 Conceptual Pipeline of the ADHD Prediction System Workflow 30
3.2 Survey Question Diagram 33
3.3 Survey Form 36
3.4 Survey Form 37
3.5 Survey Form 38
3.6 Feature Correlation Matrix 41
3.7 5-fold cross-validation 44

4.1 SHUCHOK-ADHD Risk Assessment Mobile Application (User Interface Design) 60

5.1 5 Fold Confusion Matrices for RF 62
5.2 Random Forest ROC Curve 63
5.3 5 Fold Confusion Matrices for SVM 65
5.4 ROC Curve for SVM 66
5.5 5 Fold Confusion Metrices of LR 68
5.6 ROC Curve for LR 69
5.7 5 Fold Confusion Metrics of DT 71
5.8 ROC Curve for DT 72
5.9 ROC Curve for Best Algorithm 75
5.10 Combined Confusion Matrix for Best Algorithm 76

6.1 ADHD Distribution 82
6.2 Gender Distribution by ADHD Status 83

List of Tables

2.1 Benchmarking Machine Learning Model Performance for ADHD Detection 25

3.1 Participant Summary Table 35
3.2 Feature Selection Results by SelectPercentile 42

5.1 Random Forest - Detailed Evaluation Metrics per Fold 63
5.2 SVM - Detailed Evaluation Metrices per Fold 66
5.3 Logistic Regression - Detailed Evaluation Metrices per Fold 68
5.4 Decision Tree - Detailed Evaluation Metrics per Fold 70
5.5 Comparison Summary Of All Model Performance 73
5.6 Top Correlated Features with ADHD Diagnosis 77

Chapter 1

Introduction

1.1 ADHD Overview

Attention-Deficit/Hyperactivity Disorder (ADHD) is a clear neurodevelopmental disorder that is characterized by frequent hyperactivity-impulsivity and/or inattention patterns which have a severe effect on the functioning of a person's development [1, 2]. According to the Diagnostic and Statistical Manual of Mental Disorders, the core symptoms must manifest before the age of twelve and cause clinically significant impairment in two or more settings (e.g., home, school, or social) [3, 4]. The DSM-5 [5] classifies ADHD into three presentations: predominantly inattentive, predominantly hyperactive-impulsive, and combined [6]. Inattention symptoms are associated with difficulty in staying focused on tasks or play activities. These symptoms are common and often lead to carelessness, lack of attention to instructions, and a tendency to get easily distracted. People with the predominantly inattentive presentation have poor organization of work and activity and appear to not hear when addressed [1]. Such challenges may have a severe negative impact on academic achievements of children, because they render them hard in completing homework, concentrating in school and processing information effectively [7]. Social interaction outside of the classroom can be affected by inattention by making it difficult to learn how to discuss or how to engage in group situations [3]. Symptoms of hyperactivity-impulsivity consist of rushing and overactivity. In any situation in which one is supposed to remain seated, hyperactive individuals might squirm, fidget with their seats or stand [1]. They often over-chatter, run or climb the inappropriate objects and cannot wait till their turn comes [3]. When someone makes rash decisions, they can say something without thinking, can cut-off others, or do something unsafe by the potential results [1]. These behaviors may lead to disruptive classroom behaviors, relationships among peers and an increased risk of accidents or punitive issues [7]. The underlying causes of ADHD are neurobiological. Differences are observed in brain regions associated with executive functions — working memory, inhibitory control, and cognitive flexibil-

ity. Studies have found altered activity and connectivity in the fronto-striatal-parietal-cerebellar networks that are critical for self-control and goal-directed behaviour [8–10] . These are some of the causes of the behavioral and cognitive problems observed.
The long-term effects of untreated ADHD are significant, often leading to persistent academic difficulties, professional setbacks, and an increased risk of developing comorbid mental health conditions throughout life, such as anxiety, depression, and substance use disorders. [7,11] . In order to minimize these adverse impacts, early detection and action play a vital role [12]. ADHD is a worldwide common neurodevelopmental disorder in children and teenagers. According to meta analyses, the prevalence is between 7.6% and 8.0% worldwide [13, 14]. Diagnosis rates are consistently higher in boys (˜10%) than in girls (˜5%), and this gender gap has been replicated across studies [13]. This discrepancy is in part perhaps due to girls having more internalized, inattentive symptoms that may go unnoticed, where the boys tend to display more externalized, hyperactive-impulsive symptoms which are easier to detect [3]. In Bangladesh, ADHD prevalence in children is estimated at 3-6% [3]. Sociodemographic variables, including parental education and socioeconomic status, may influence these rates and their development, as well as the prevalence of comorbid conditions, according to national research. [3]. To address the potentially large number of undiagnosed cases, this situation highlights the urgent need for accessible and customized screening technologies.

Neurobiological Differences between the Brain of an ADHD patient and a Neurotypical one: ADHD is closely related to anatomical and functional dysfunctions of particular areas of the brain involved in executive functioning and attention. The condition has been associated with poor control and inhibition of impulses, with brain areas, such as the prefrontal cortex, basal ganglia, and cerebellum, playing important roles in the decision making process, behavioral inhibition, and cognitive flexibility. Based on neuroimaging studies, people with ADHD tend to have an abnormal pattern of brain activity. The brain of ADHD is more prone to display greater variability in the signalling of neurons and reduced activity in regions of the brain that are important in sustaining attention in comparison to neuro typical individuals. Such differences lead to such common symptoms as impulsiveness, distractibility, and inability to focus on the tasks. Figure 1.1 shows a comparison of ADHD and non-ADHD brains using fMRI, showing differences in executive control and activity regulation. Note that the figure is illustrative only and is not intended as diagnostic instrument.

ADHD Prevalence Worldwide: ADHD is one of the prevalent neurodevelopmental disorders in children and adolescents across the globe. Numerous epidemiological research and meta-analyses have provided estimates of its prevalence in children worldwide with estimates ranging between 7-8%. But, because of the differences in the diagnostic practices, prevalence rates vary significantly across regions, depending on cultural attitudes, awareness levels and access to healthcare services.

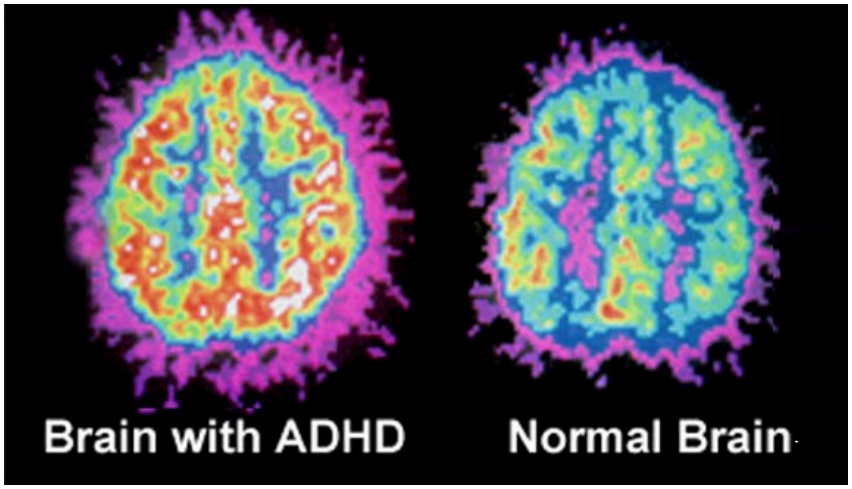

Figure 1.1: ADHD Brain Vs Normal Brain [Source]

In developing countries, underdiagnosis is more prevalent, and in developed countries, higher rates of diagnosis tend to be reported, due to the presence of better screening programs and awareness. Lack of access to diagnostic instruments, social stigma and lack of knowledge contribute to the huge number of undiagnosed cases in South Asian countries like Bangladesh. This gap shows the dire necessity of scalable, low-cost screening technologies.

Figure 1.2 provides an overview of worldwide ADHD prevalence and shows regional variations.

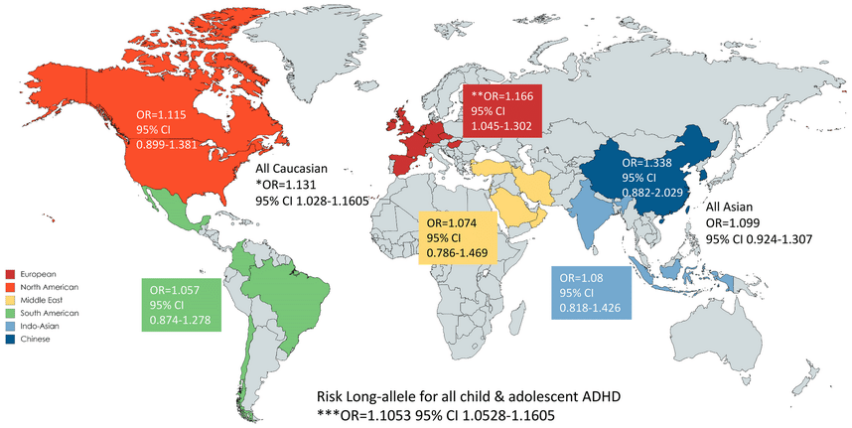

Figure 1.2: Worldwide ADHD Overview [Source]

1.2 Problem Statement

Despite the known prevalence of ADHD in Bangladesh, there is a significant gap between expected cases and children who receive an early diagnosis and effective treatment. Several difficulties are interconnected and cause this lack of diagnosis and treatment in the

Bangladeshi setting. Lack of awareness among parents, teachers, and even some medical professionals is a major contributing factor, particularly regarding the symptoms of ADHD and their impact on development [15]. There is also a significant barrier to accessing treatment, namely the social stigma surrounding mental health conditions such as ADHD [16]. In most cultures, mental health issues are stigmatized hence causing families to hide symptoms or steer clear of professional evaluation to avoid prejudice or social rejection [17]. This stigma is not unique to low- and middle-income countries, it is also a significant cause of treatment gaps in higher-income countries [18]. For most Bangladeshi families, the cost of conventional ADHD evaluation - involving in-depth clinical ratings, multi-informant assessments and specialised tests - is prohibitive [19]. Because of this cost and unqualified mental health practitioners and diagnostic centers, particularly in rural areas, a significant portion of the population cannot obtain the conventional diagnostic pathways. [20–22] . This means that a large number of cases of ADHD are not treated and this may have adverse long term consequences such as academic inefficiency, disruptive behavior and delayed social-emotional development [7]. This study meets this requirement by creating a culturally-appropriate, data-driven, and simple machine-learning application to screen ADHD symptoms among Bangladeshi children. This study will suggest an easy to use and affordable way of early detection in an environment where standard diagnostic services are scarce. The approach helps to make timely intervention and better outcomes by using non-clinical and survey-based data.

1.3 Objectives

This study aims to enhance the prediction of ADHD among Bangladeshi children using machine learning with the goal of supporting early diagnosis and intervention. The specific objectives of this work are as follows:

- To identify the most significant behavioural and environmental factors that act as predictors of ADHD in the Bangladeshi paediatric population.
- To determine the best suitable model in predicting ADHD in given dataset, an exhaustive comparative study will be carried out using Random Forest, Support Vector machine (SVM), Logistic Regression, and Decision Tree classifiers.
- To suggest the creation of a user-friendly, multilingual (English and Bengali) online interface, which can be used as a self-assessment and awareness tool, allowing parents, teachers, or guardians to get an approximate risk of ADHD based on the developed machine learning model.

1.4 Key Contribution of This Study

The study contributes to the knowledge on machine learning-based medical solutions and ADHD screening. particularly in low resource areas such as Bangladesh:

- Translating DSM-5 based questionnaires into Bengali will allow the design of a culturally sensitive ADHD screening instrument, which will result in a higher level of contextual relevance, higher levels of response accuracy and greater validity of data collected.
- Non-Clinical, Low-Cost Prediction Model: This study proposes a screening that uses machine learning models that uses survey information can be easily obtained, thus early detection is more cost effective scalable, compared to traditional means of diagnostic testing that rely on neuroimaging or clinical expertise.
- Comparative Analysis of Machine Learning Models: To determine an optimal approach to predict ADHD in the given data, a detailed evaluation of the Random Forest, Support Vector Machine, Logistic Regression and Decision Tree models are performed.
- The key behavioral predictors, including impulsivity, fidgeting, distractibility, and inability to remain focus during studying are found to be strong predictors of ADHD among Bangladeshi children.
- Prototype of a Bilingual Mobile-Based Screening Tool: To assist parents, educators, and caregivers with self-assessment and guidance, an easy-to-use bilingual (English–Bengali) interface application is proposed.
- Concentrate on Bangladesh-Specific Solution: By providing a solution that is specific to the Bangladesh socioeconomic and cultural conditions, this study addresses the lack of locally based strategies of ADHD screening.

  Combined, these initiatives aim to improve early detection of ADHD, awareness creation, and early intervention in impoverished communities.

1.5 Research Questions

This study is guided by two primary research questions, aimed at identifying key factors in the Bangladeshi context and explore the efficiency of machine learning in predicting ADHD:

- RQ1: Can machine-learning models trained on non-clinical, survey-based data reliably predict ADHD risk in Bangladeshi children?

- RQ2: What are the most significant environmental and behavioural predictors of ADHD in Bangladeshi children, and how do they affect the diagnostic accuracy of machine-learning models?

1.6 Limitations and Challenges

The limited sample size of our data (N = 110) may constrain the generalisability of the findings and raise the risk of overfitting — i.e., models may perform well on the training data but poorly on unseen cases [23–25]. Because the dataset is specifically culturally sensitive to Bangladesh, the models may not generalise to other populations without further validation, raising concerns about external validity [26, 27] . Though survey-based information is feasible in terms of convenience, it has a risk of bias impacting data integrity e.g. social desirability or recollection errors [28, 29] . Furthermore, a dichotomous conceptualization of ADHD oversimplifies a neurodevelopmental disorder, potentially missing the conceptualization of differences in symptom presentation and severity. This constraint may frustrate an in-depth study of the disorder [30, 31]. All these limitations point out to the areas that should be further investigated and improved to increase the robustness, resilience, and wider applicability of the developed prediction model.

Chapter 2

Literature Review

2.1 Materials and Methods

To identify and synthesize relevant studies on machine learning-based prediction of ADHD, an extensive and systematic search of the literature was conducted. The general goal of this search strategy was to gather a list of a large number of peer-reviewed articles explaining various machine learning methods, their applications, and their applicability in the diagnosis or prediction of ADHD. Key academic electronic databases were systematically searched to ensure that all relevant fields were covered. They were PubMed, Scopus, Web of Science, and Google Scholar, all of which offered significant benefits within computer science, psychology, medicine, and neurology.

To identify articles within the nexus of ADHD, machine learning, and children, search terms were carefully designed with a mixture of relevant keywords. Keywords and phrases comprised ADHD prediction, machine learning, artificial intelligence, diagnosis, children, adolescents, behavioral features, demographic factors, survey data, and related terms: neurodevelopmental disorders and the corresponding developing countries. These phrases were mixed carefully with the use of Boolean operators such as AND to narrow down a search (e.g., AND (machine learning OR artificial intelligence) AND (prediction OR diagnosis)) and OR to expand them (e.g., children OR adolescents). This repeated keyword refining technique was aimed at reducing the number of irrelevant results and maximizing the retrieval of relevant content.

The primary focus was on articles published within the last decade to ensure the timeliness and up-to-date relevance of the review due to the rapid changes in machine learning methods. Nevertheless, older works, though seminal, that were the basis of groundwork or offered critical context were also taken into account. The articles that were found were then subject to a multi-phase screening process after the initial searches of databases. To begin with, titles and abstracts were thoroughly reviewed to be directly related to the purpose of the project. Preference was given to empirical studies that

clearly employed machine learning methods to diagnose or predict ADHD, particularly those that employed behavioral or demographic factors in pediatric samples. Studies that merely used neuroimaging without a behavioral connection, genetic indicators, or overall treatment effectiveness were typically filtered out unless they provided significant methodological contributions to the use of ML. The complete-text articles of all the potentially relevant studies were then accessed and thoroughly evaluated according to a list of predefined inclusion criteria: the research should be peer-reviewed, written in English, and include original research results with sufficient methodological details. This methodical and open approach ensured the choice of good and directly relevant literature to inform the next discussion of related work and its findings.

The Preferred Reporting Items of Systematic Reviews and Meta-Analyses (PRISMA) was followed to ensure transparency and reproducibility of the study selection process. A systematic search, screening, and selection of relevant studies to achieve this review are depicted in Figure2.1. To begin with, the comprehensive search in multiple databases yielded a great number of results. The rest of the research was then filtered out according to the titles and abstracts to identify which ones could be used once the duplication was removed. Pre-existing inclusion and exclusion criteria were then used to evaluate full-text publications. The best and most relevant studies were incorporated in the final review, and those which failed to meet the requirements were filtered out.

2.2 Related Work

Machine learning methods have significantly influenced the diagnostic and predictive tool of ADHD, with potential answers to the resource-intensive and subjective quality of the traditional clinical assessment [22, 32]. What is fueling the shift to ML-based methods is the possibility of better impartiality, scalability, and accessibility, particularly in cases where there are limited expert clinical resources.

Early Methods: Previously the initial machine learning models used to predict ADHD relied heavily on neuroimaging data, such as electroencephalography and functional magnetic resonance imaging (fMRI), and aimed at predicting distinct neurobiological signs of the disease [33, 34]. Indicatively, the research studies which examined fMRI data with Support Vector Machines and Artificial Neural Networks could differentiate persons with ADHD and neurotypical controls with 70-90% accuracy [35]. Specifically, research has been conducted with the Fractional Amplitude Altered Low-Frequency Fluctuation along with the Relief algorithm and SVM on neuroimaging data with high classification accuracies, including 98.04% [36]. Similarly, Omneya Attallah introduced an EEG-based tool, "ADHD-AID," which used multi-resolution analytic techniques to extract various nonlinear, band-power, entropy-based, and statistical features that gave it an exceptional accuracy (0.991), sensitivity (0.989) and specificity (0.992) [37]. Another

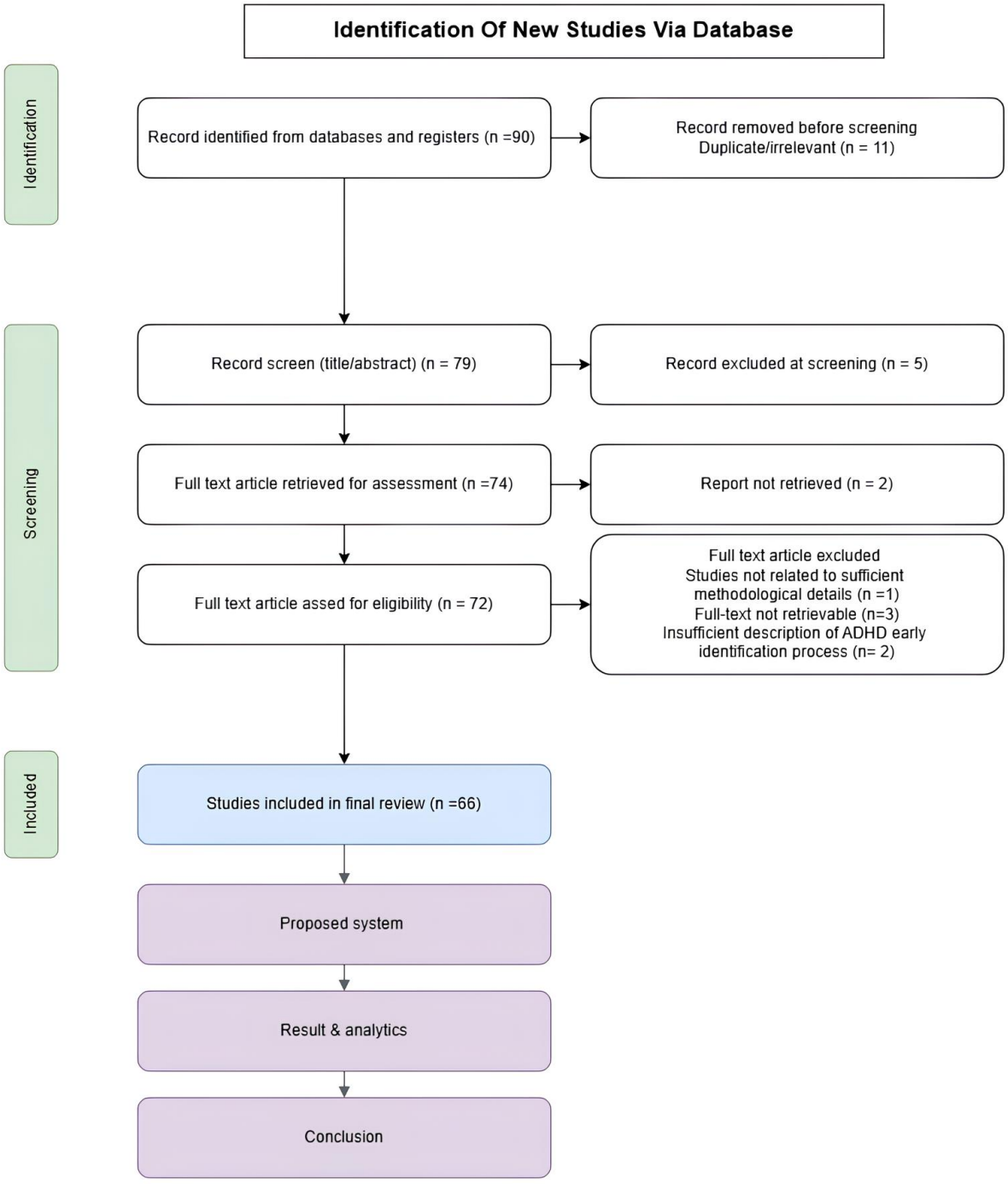

Figure 2.1: PRISMA flow diagram

web-based application in a case study (DETEC-ADHD) performed logistic regression analysis on clinical and EEG data, which yielded 90% accuracy in adults and 100% in children. This implies that EEG can be used concurrently with psychological tests to enhance quality of diagnosis to a significant degree [36]. However, despite their high predictive potential, these techniques based on neuroimaging are expensive and require specific equipment, which often limits their usage in settings with limited resources to avoid the necessity of alternative data types that are available easily [32, 33].

Transition to Demographic, Clinical, and Behavioural Data : Recent studies have also extensively explored machine learning applications in practice using more readily available data, including behavioral tests, demographic data, and clinical history, in acknowledgment of the neuroimaging constraints. This is in close relation to the objectives of the project which focuses on non-clinical, survey-based data. Mooney et al. state that detailed evaluations of ADHD require a significant amount of resources such as standardized ratings by multiple informants and a systematic clinical interview, which often are beyond the capacity of primary-care clinicians [32]. Their research investigated the application of fast, cheap clinical measurements, based on data types that are widely used in typical clinical practice, and found that the measurements were able to forecast the diagnosis of ADHD, potentially recreating gold-standard diagnostics [32].

Behavioral symptom assessments have been successfully applied using machine learning in a number of studies. Random Forest-based classifier using behavioral activity data of children aged 3 to 17 gave a high accuracy of 85.5, sensitivity of 84.4, specificity of 86.4 and an AUC of 0.94, which means that at least RF and Logistic Regression can give outstanding results in terms of accuracy [38]. Slobodin et al. demonstrated high accuracy (87%), sensitivity (89%), and specificity (84%) of CPT data, which was 34% worse than benchmark models [39]. On the basis of demographic data, the results of Behavioral Rating Scale and Wechsler Intelligence Scale, Qin et al. developed interpretable machine learning models to identify ADHD in children. With binary classification, their RF model achieved an AUC of 0.99 and an accuracy of 0.95; with ADHD subtype classification, an SVM model achieved a micro-average AUC of 0.96 and an accuracy of 0.84. The models are practical and scalable as they can be easily acquired at the normal clinical visits [40].

The Performance of Common Machine Learning Models : The choice of the ML algorithm is one of the key factors influencing predictive performance. Several algorithms are commonly used:

- Random Forest: This is among the most common top performers due to its resistance to overfitting, handling of diverse and high-dimensional data, and the ability to perform meaningful analysis of feature importance. Research shows high accuracy, precision and recall when applied, such as Maniruzzaman et al.'s 85.5% accuracy [38] and Qin et al.'s 95% accuracy and AUC of 0.99 [40].

- Logistic Regression: These are often used due to their effectiveness for classification problems, particularly for two classes with clear decision boundaries. LRs can have high micro-average AUCs for multiclass problems, and competitive accuracy rates, often over 80%, particularly when combined with good features [36, 40] .

- Decision Trees: They offer unambiguous rule-based decision routes and are valued for their interpretability and simplicity. While DTs can overfit, they are used as the building blocks for other powerful, ensemble models such as Random Forest. For instance, Chen et al. obtained 85.507% accuracy with a Decision Tree model using self-reported questionnaires and demographic information [36].

- Deep Learning: Advanced techniques such as Convolutional Neural Networks (CNN) and Multi-Layer Perceptrons (MLP) are increasingly applied, most often with large data sets, and often have very high accuracy with sufficient data [38]. Such as de Lacy and Ramshaw used multimodal predictors from a large population and deep learning to predict ADHD [41]. Garcia-Argibay et al. employed deep learning to predict the onset of ADHD in a Swedish national registry data of children and adolescents [42].

Problems and Issues with ML-Based ADHD Prediction Despite these developments, the profession still faces several difficulties. Small sample size is a typical limitation in research that may limit the generalizability of findings and increase the risk of overfitting, where models work very well with the data they were trained on but perform poorly with new data [24, 32]. Mengi and Malhotra observed that the majority of studies on the diagnosis of socio-behavioral disorders have relatively small sample sizes (less than 200) that directly impact the effectiveness and the generalizability of the diagnostic models [24]. This is particularly important for complex machine learning models that require large data sets to be trained properly without overfitting [32].

Another major issue is the generalizability of the ML models across different demographics and cultural contexts. Models that work well in Western populations may not be as effective in other cultures due to different symptom patterns, diagnostic tests and socioeconomic factors [43]. For instance, Brown et al. found models that only included demographic data performed slightly better than models that included fMRI features (62.5% accuracy), but results can be influenced by sample variability and unbalanced data, which is common in real-world clinical datasets [43]. To ensure fairness and generalizability, Ter Minassian et al. highlighted the need to use re-weighting algorithms to remove biases in clinical data [44]. This underlines the need to use culturally sensitive tools and collect data in the area of interest, such as in Bangladesh, as this study does, to ensure the model's accuracy and effectiveness [22].

In addition, there is a need for "black-box" or complex machine learning models to be clinically acceptable for their interpretability. Clinicians need to understand the

rationale behind a prediction to trust and use a model [40]. Methods such as Shapley Additive explanations are increasingly being applied to enhance model interpretability and trustworthiness by providing information about the features [40]. Finally, data quality challenges that require substantial preprocessing and validation are presented by methodological issues, like missing data and potential biases in self-reported or survey data (e.g., recall bias, social desirability bias) [32, 45]. The current project aims to address these challenges by focusing on a data-driven, culturally sensitive approach, tailored to the Bangladeshi context.

2.3 Literature Search Outcome

Several recurring traits and patterns regarding the use of machine learning in ADHD prediction were found by the thorough literature search and subsequent analysis of pertinent studies. These insights are essential for determining the current project's distinctive contributions and placing it within the body of previous knowledge.

Population Demographics and Geographical Scope: The ages of the children and adolescents involved in most of the studies identified were typically from early childhood (e.g., 6 years) to adolescence (18 years), which is consistent with the age of onset of ADHD [46, 47]. Some studies also included young adults, particularly for the purpose of studying ADHD prognosis. In terms of geographical representation, studies from Western developed countries (especially the United States, Canada and Europe) were clearly over-represented [44, 46]. Given that there were fewer studies mainly from South Asia or other developing countries that examine ADHD prediction using ML in a local environment, this geographical gap highlights an important research gap. To ensure cultural relevance and generalizability of prediction models to diverse populations, this lack of research emphasises the need for localised research, such as the present study. For instance, although some studies referred to issues in developing countries in general [22], there is still a lack of practical applications of machine learning using behavioural data, especially in South Asian settings like Bangladesh.

Methodological Approaches and Data Characteristics: While there are multiple methods of predicting ADHD using ML, some patterns exist. A variety of data was used, which might be broadly classified into:

- Neuroimaging Data: fMRI, EEG, and structural MRI data were widely used in previous and some current studies to identify neuroimaging-based biomarkers [33] [34] [36].

- Behavioral and Clinical Data: Use of behavioral rating scales (parent, teacher), clinical interview, neuropsychological test measures and direct observations of behavior is an emerging practice that is particularly relevant to our study [32] [40].

- Demographic and Administrative Data: Many studies used large-scale health and education administrative data combined with demographic data (age, gender and socioeconomic status) [44] [46].

Commonly used features in such data sources were measures of inattention (e.g., inattention, distractibility), hyperactivity (e.g., fidgetiness, hyperactivity), impulsivity (e.g., impulsivity, interruption), various executive dysfunctions, and basic demographic variables. A key development is adopting more accessible behavioral and demographic data, driven by a goal of creating cost-effective screening tools that can be used beyond specialised clinical settings [32]. Unlike neuroimaging, Han Qin et al. in particular, noted that the demographic, behavioral rating scales and IQ tests are accessible, and hence suitable for practical ML use [40].

Performance Metrics and Model Efficacy: Well-established classification metrics were found to be regularly used to evaluate model performance. These included accuracy, sensitivity (recall), specificity, precision, F1-score and the Area Under the Receiver Operating Characteristic Curve (AUC). The ability of ML models to discriminate between ADHD patients and controls was shown by many studies that lead to high AUC values, frequently greater than 0.90 and often greater than 0.80. For example Random Forest models have consistently demonstrated their prediction ability through the attainment of accuracies greater than 85% and AUCs greater than 0.90 [38] [40]. But cautionary notes regarding potential biases introduced by sample size, class distribution or the use of the specific ADHD diagnostic criteria were often included as well as the interpretation of these scores.

Recurring Limitations and Gaps: There were a number of common limitations highlighted in the literature:

- Limited Sample Sizes: Many machine learning studies on predicting ADHD have small sample sizes, limiting the statistical power and generalizability of the findings, particularly those using advanced data such as neuroimaging [24] [32]. This increases the risk that models might work well with the data used to train them, but not when applied to new data.

- Limited External Validation: While the rigorous external validation of models on independent data is less common than cross-validation within the same data set, this suggests many claimed models may not be generalizable [32].

- Diagnostic and Other Biases: Direct comparisons and meta-analyses may be confounded by heterogeneity and potential biases due to diagnostic criteria (e.g., DSM-IV vs. DSM-5), measurement tools and therapeutic procedures across studies [45]. Selection bias is a bias that has been well-studied in diagnostic test accuracy studies.

- Cultural Relevance and Transferability: Most models that are currently being used have been developed and validated in Western populations. They may not be readily applicable to non-Western populations without substantial adaptation and verification due to cultural variations in the presentation and reporting of symptoms, access to health care and cultural attitudes [43].

- Interpretability: More complicated machine learning models, such as deep learning models, are often "black boxes" that are hard for clinicians to understand the rationale behind a prediction, and therefore may not be implemented in clinical practice [40].

As well as identifying the important gaps that the current study (which is on culturally sensitive survey-based ML in Bangladeshi children) aims to address, this overview of the literature search outcomes illustrates the significant progress made in ML-based ADHD prediction.

2.4 ML Model Popularity

Machine learning has allowed many algorithms to be popularly used in ADHD studies, each with features that have different advantages for tackling the difficulties of ADHD prediction. The popularity of these models is often due to their interpretability, robustness or handling of data.

Random Forest (RF) models are frequently used due to their many advantages. They are not prone to overfitting, can deal with highly variable and high-dimensional data, and provide high feature analysis and importance assessment [38]. RF models are also found to be among the best in many studies in terms of recall, accuracy and precision. For example, a Random Forest model achieved an impressive AUC of 0.99 and accuracy of 0.95 for binary ADHD classification in a study that used demographic information, Behavioral Rating Scale (BRS) data and Wechsler Intelligence Scale (WISC) scores [40]. Another study using data on activity levels also reported an RF-based classifier with an accuracy of 85.5% [38].

Another popular alternative is **Support Vector Machines (SVM)** that are highly valued for their speed in classification problems when a clear separation between classes is needed. With the right feature set, SVMs have demonstrated competitive accuracies that often top 80%. They also have shown quite high micro-average AUCs in multi-class tasks, such as ADHD subtyping [40]. SVMs were broadly used in early studies of neuroimaging due to their ability to model complex brain data [35] [36].

Despite being a simple statistical model, **Logistic Regression (LR)** remains popular because of its interpretability. It offers insight into how predictions are made by capturing the relationship between an outcome (such as having ADHD) and predictors. When

datasets are small and the data is "well-behaved", simple LR models can perform better than complex algorithms [48]. For example, in a case study, a web tool (DETEC-ADHD) used logistic regression and achieved an accuracy of 90% for adults and 100% for children [36]. Another study found that the LR was able to predict ADHD in population samples with an AUC of 0.86 [44].

Deep Learning models, such as Convolutional Neural Networks and Multi-Layer Perceptrons are increasingly being used, especially with the availability of larger and more complex data. They are often highly accurate when training data is available as they are very good at learning non-linear patterns [38]. Large-scale studies have successfully applied deep learning to leverage multimodal features to predict the development of ADHD [41].

Decision trees (DT) are popular for their simplicity and the different rules that can be used to reach a decision. While individual DTs are known to overfit, their clinical utility in difficult classification tasks is enhanced as they are the building blocks of powerful ensemble methods such as Random Forest [36].

A trade-off between the characteristics of the data, the need for model interpretability and the specific goals to be predicted is usually required when selecting a machine learning model. In clinical practice, the latest trends are towards models that are both useful and predictive.

2.5 Heatmap and Model Usage

A conceptual heatmap would be great to show the popularity and success of machine learning models in ADHD-related research. Such a heatmap could potentially illustrate the intersection of a number of machine learning algorithms with a variety of data types and, perhaps, their achieved accuracy. The heatmap's intensity could depict the extent to which a particular model is applied to a certain kind of data or the reported effectiveness of the model.

The following is an illustration of what the heatmap would show:

- **Neuroimaging Data vs. Models:** Regions associated with neuroimaging data (e.g., fMRI, EEG) would probably have high intensity with Support Vector Machines and certain Deep Learning models. This is because they are especially adept at handling the high dimensionality and non-linearity in neural data, as illustrated in studies that use EEG for ADHD diagnosis [37] and fMRI [35]. The "high density" would indicate they are used often and successfully.

- **Behavioral, Clinical, and Demographic Data vs. Models:** For more readily available data sources such as behavior rating scales, clinical measures and demographic data, the "heat map" would display a high density for Random Forest,

Support Vector Machines, and Logistic Regression. These models have been applied with great success to these types of data for the prediction of ADHD (RF for its high accuracy) [38] [40], LR for its interpretability and good results [44]. Their popularity could be explained by their use of easily accessible, non-invasive data.

- **Large-Scale Data vs. Models:** New research that uses very large data sets (such as national registries or large population based data sets) would have an increasing density for Deep Learning models [41] [42] Their capacity to learn complex patterns in large data sets renders them well suited for such predictive tasks.

- **Interpretability vs. Performance Trade-off:** Implicitly, the heatmap could also show a trade-off. Models easier to understand (such as Logistic Regression) and Decision Trees would be more likely to be used in areas where clinical understanding of the prediction process is crucial. On the other hand, highly predictive but "black-box" models such as deep neural networks would have higher performance measures in some domains but would be less accepted by clinicians due to their opacity [40].

The shift from predominantly neurophysiological data to more accessible behavioural and demographical data, and the corresponding shift in favoured machine learning techniques, would effectively illustrate the proposed heatmap in Figure 2.2.

2.6 Benchmarking Model Performance

Benchmarking the performance of machine learning models can be used to provide a critical comparative context, allowing an objective assessment of different algorithmic methods in ADHD prediction. The key performance measures of previous studies are summarized in Table 2.1 below and offer a comprehensive guide on the expected success of various strategies. Along with the information about the sample size, and some data characteristics, it includes such metrics as Accuracy, Precision, Recall, Specificity, F1-Score, and Area Under the Curve, where they are provided in the provided context. A measure is denoted by N/A when it was not reported in the text to a particular study.

Table 2.1: Benchmarking Machine Learning Model Performance for ADHD Detection

Study	Model	Type	Data	N	Acc	Prec	Rec	Spec	F1	AUC	Notes
Attallah et al. [37]	ADHD-AID	EEG	EEG features	N/A	0.991	N/A	0.989	0.992	0.989	0.998	EEG-based ADHD detection
Attallah et al. [37]	M-SVM	EEG	EEG features	N/A	N/A	N/A	N/A	N/A	N/A	0.998	Strong classification
Attallah et al. [37]	k-NN	EEG	EEG features	N/A	N/A	N/A	N/A	N/A	N/A	0.997	High performance
Attallah et al. [37]	C-SVM	EEG	EEG features	N/A	N/A	N/A	N/A	N/A	N/A	0.996	Exceptional results
Attallah et al. [37]	Q-SVM	EEG	EEG features	N/A	N/A	N/A	N/A	N/A	N/A	0.994	Strong performance
Attallah et al. [37]	ANN	EEG	EEG features	N/A	N/A	N/A	N/A	N/A	N/A	0.990	Strong performance
Santarrosa-López et al. [36]	SVM	Neuro	fALFF	N/A	0.980	N/A	N/A	N/A	N/A	N/A	Neuroimaging features
Santarrosa-López et al. [36]	LogReg	Clinical	EEG + Clin.	N/A	0.90–1.00	N/A	0.923	N/A	N/A	0.966	DETEC-ADHD system
Maniruzzaman et al. [38]	RF	Behav.	Activity	N/A	0.855	N/A	0.844	0.864	N/A	0.940	RF improves accuracy
Slobodin et al. [39]	ML	General	CPT data	N/A	N/A	N/A	N/A	N/A	N/A	N/A	34% improvement
Qin et al. [40]	RF	Clinical	Demo + WISC	N/A	0.95	N/A	0.95	0.95	0.95	0.99	Strong clinical model
Qin et al. [40]	SVM	Clinical	Demo + WISC	N/A	0.84	N/A	N/A	N/A	0.84	0.96	Multiclass classification
Ter-Minassian et al. [44]	RF	Linked	Edu + Health	N/A	N/A	N/A	N/A	N/A	N/A	0.86	High predictive acc.
Ter-Minassian et al. [44]	LogReg	Linked	Edu + Health	N/A	N/A	N/A	N/A	N/A	N/A	0.86	High predictive acc.
López et al. [35]	SVM	EEG	EEG	N/A	0.949	N/A	N/A	N/A	N/A	N/A	SVM widely used
Brown et al. [43]	LogReg	Demo	Demographic	N/A	0.625	N/A	N/A	N/A	N/A	N/A	Simple data effective
Colby et al. [43]	ML	General	Demographic	N/A	0.627	N/A	N/A	N/A	N/A	N/A	Outperformed MRI
Chen et al. [49]	Dec. Tree	Clinical	Survey	69	0.85507	-	-	-	-	-	Uses accessible data.
Chen et al. [49]	Hybrid	Clinical	Study	N/A	0.9361	-	-	-	-	-	Hybrid ADHD model.
Chen et al. [49]	ML Model	Clinical	Study	N/A	0.7503	-	-	-	-	-	All features used.
Chen et al. [49]	Rule-based	Clinical	Study	N/A	0.6527	-	-	-	-	-	Rule-based ML.
Mooney et al. [48]	Bayesian	Survey	Parent	>1000	>0.86	-	-	-	-	-	Predicts diagnosis.
Garcia-Argibay [42]	DNN	Registry	Sweden	238696	0.69	-	0.7166	0.650	-	0.75	Balanced accuracy.
de Lacy et al. [41]	Multimodal	ML	ABCD	N/A	0.86–97	0.82–97	0.82–97	-	-	0.91–99	Predicts ADHD/ODD.

Continued on next page

Table 2.1 – continued from previous page

Study	Model	Type	Data	N	Acc	Prec	Rec	Spec	F1	AUC	Notes
Liu et al. [46]	ML	Health	Admin	23494	-	-	-	-	-	0.811	Reliable prediction.
Chen et al. [49]	SVM	Clinical	Assess.	69	0.72464	-	-	-	-	0.784	Based on 94 vars.
Chen et al. [49]	LogReg	Clinical	Assess.	69	0.72464	-	-	-	-	0.795	Based on 94 vars.
Chen et al. [49]	Dec. Tree	Clinical	Assess.	69	0.82609	-	-	-	-	0.866	Strong performance.
Chen et al. [49]	KNN	Clinical	Assess.	69	0.59420	-	-	-	-	0.558	Lower performance.
Chen et al. [49]	RF	Clinical	Assess.	69	0.81159	-	-	-	-	0.866	Good performance.
Chen et al. [49]	Naive Bayes	Clinical	Assess.	69	0.75362	-	-	-	-	0.870	Stable results.
Alsharif et al. [50]	SVM	Dataset	Standard	N/A	0.91	-	-	-	-	-	Highest accuracy.
Alsharif et al. [51]	AB	Text	Reddit	N/A	0.79	0.82	0.82	-	0.82	0.78	Text classification.
Alsharif et al. [51]	GRU	Text	Reddit	N/A	0.78	0.79	0.85	-	0.82	0.77	Good recall.
Alsharif et al. [51]	LSTM	Text	Reddit	N/A	0.77	0.81	0.81	-	0.81	0.77	Balanced results.
Alsharif et al. [51]	XGBoost	Text	Reddit	N/A	0.81	0.83	0.84	-	0.84	0.80	Strong results.
Alsharif et al. [51]	RF	Text	Reddit	N/A	0.811	0.843	0.837	-	0.840	0.81	Best among models.

The diversity of machine learning applications to ADHD prediction is illustrated by this large table. Models trained on neuroimaging information (e.g. EEG) tend to report very high performance metrics. An example is the ADHD-AID tool developed by Attallah [37], which achieved an accuracy of 0.991 and high sensitivity and specificity with the help of some EEG variables. ADHD-AID models (M-SVM and k-NN) exhibited AUCs of approximately 0.997. Likewise, DETEC-ADHD by Santarrosa-Lopez et al [35], With logistic regression to combine clinical and EEG data, very high accuracies were found, including in a case study 90 -100 and a recall of 0.923 with an AUC of 0.966 with a subset of features. Such results show the high predictive power of neurophysiological factors. Random Forest and Support Vector Machine models regularly demonstrate good performance with more readily available data, such as clinical assessments, behavioral observations and demographic information. In Maniruzzaman and others [38] based on the data of behavioral activities, a Random Forest model was created with an impressive result of 85.5% accuracy, a recall of 0.844 and a specificity of 0.864 and an AUC of 0.94. Qin and associates [40] disclosed an SVM model of multiclass classification with a F1-score of 0.84 and an AUC of 0.96, and a Random Forest model with a 95% accuracy, 0.95 F1-score and an AUC of 0.99 using demographic and behavioral rating scale data. Even with an AUC of 0.86 in large population studies by Ter-Minassian et al., even logistic regression, often considered a baseline, has large predictive power [44]. Interestingly,

other studies, such as those by Brown et al. and Colby et al. [43], demonstrated that in some competitive situations, the use of simple demographic data, as opposed to a more complex neuroimaging data, might sometimes be more effective, which underscores the significance of readily available data. National registry data studies reported balanced accuracy = 0.69 and AUC = 0.75 [42]. Multimodal models by de Lacy et al. [41] achieved accuracy values ranging from 0.86–0.97 and AUC values ranging from 0.919–0.996. Liu et al. [46] reported an AUC of 0.811 using administrative health data. Chen et al. [49] and Alsharif et al. [50] utilize diverse data sources, including clinical evaluation data or text features of posts on Reddit, to explore more in-depth a range of machine learning models, including SVM, Logistic Regression, Decision Tree, K Nearest Neighbor, Random Forest, Naive Bayes, AB Model, GRU, LSTM, and XGBoost. There are numerous accuracies and AUCs shown by these experiments, and it is reasonable to assume that the choice of the algorithm, data source, and feature set influence the model performance considerably. An example of this is a Decision Tree model of Chen et al. [49], which had an accuracy of 0.82609 and an AUC of 0.866, and Naive Bayes had 0.870 AUC. Alsharif et al. [50] demonstrated that the Random Forest has a promising performance of 0.811 accuracy and 0.81 AUC on the text features of Reddit. In a clinical study, Chen et al. [49], also showed a hybrid model with 0.9361 accuracy. This critical evaluation shows that even though neuroimaging-based methods can achieve really high accuracy, convenient and workable methods utilizing behavioral and demographic data and models such as Random Forest and SVM can achieve a lot of predictive power. To create scalable and efficient ADHD screening systems, it is important to create a balance between high performance and access to the data, especially when working with resources-limited settings. According to the limitations of the field more generally, there is still no easy way of ensuring that these high performance indicators cut across a range of larger, externally validated datasets [32].

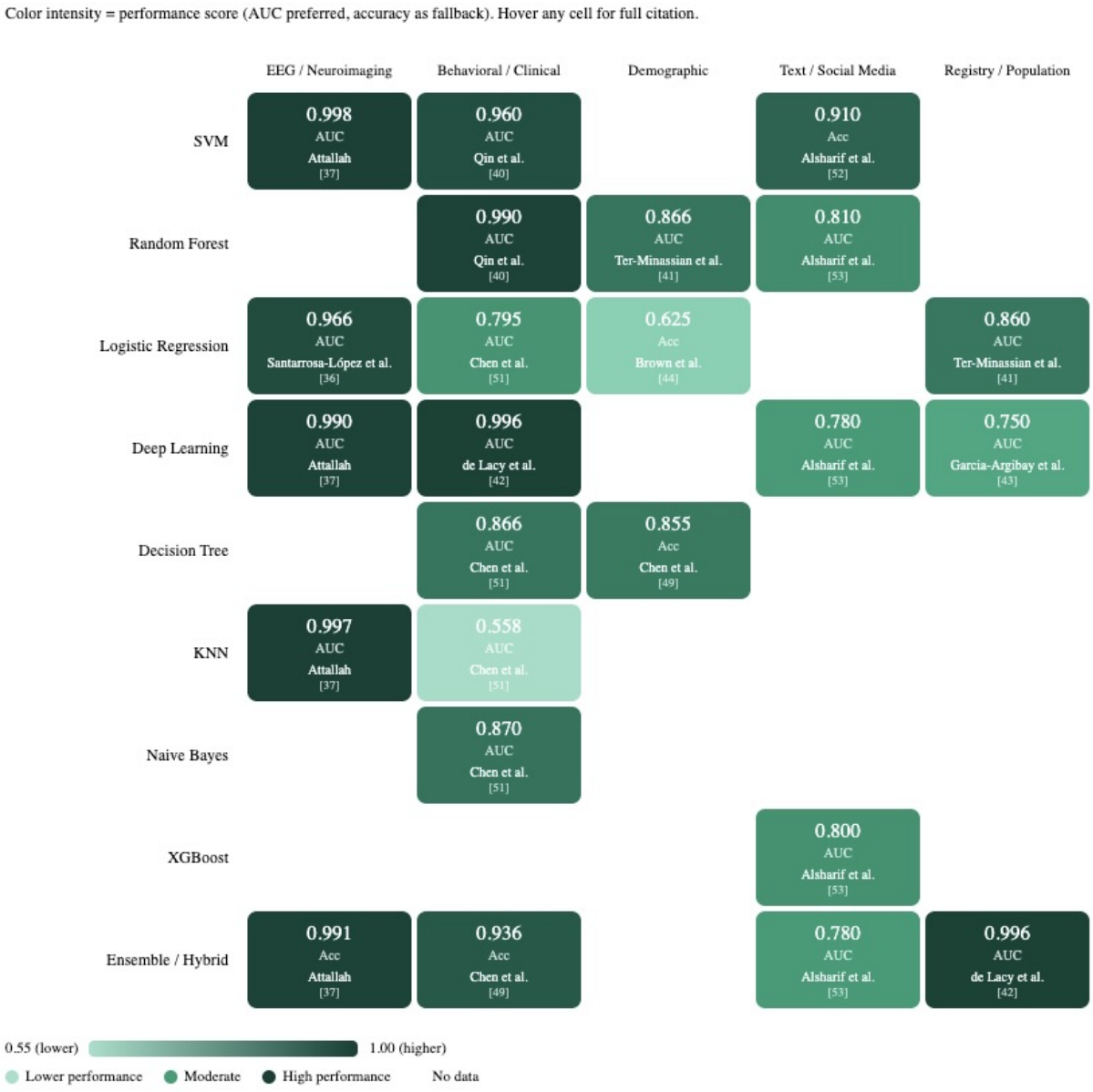

Figure 2.2: Conceptual Heatmap

Chapter 3

Methodology

This section explains the systematic approach taken to design and test a machine learning-based model to detect ADHD in Bangladeshi children at the early stages. It encompasses all the stages of the process. The approach employs powerful machine learning algorithms and survey data which has been culturally optimized to the study issues.

3.1 System Architecture

The general system architecture of this project is imagined to be a pipeline of efficiently engineered stages that are interconnected and sequential in order to transform the raw survey data into useful bits of information to assess ADHD risk. To develop a conceptual interface that is easy to use in the real world, this comprehensive workflow in Figure 3.1 incorporates several essential steps. The first step of the process is data collection which is the primary way of collecting data. In this case, a non-clinical, culturally appropriate questionnaire is employed to collect unobtrusive behavioral and demographic information that are directly connected with the ADHD symptomology in Bangladesh. This step is highly crucial as the validity and effectiveness of all other stages of analysis and prediction depend directly on the intrinsic quality, cultural sensitivity, and representativeness of the collected data. It is the non-clinical design of the survey that is expected to fill a significant gap in the early detection capabilities in resource-limited settings by defeating conventional diagnostic limitations such as cost and availability [33].

Data Preprocessing stage follows the data gathering stage and involves a lengthy process on the raw survey responses. This step is necessary to transform potentially inconsistent, noisy, and raw data into a clean, structured and machine readable form that can be fed into machine learning algorithms. Important preprocessing techniques include whitespace normalization (to ensure consistency in text-based entries) and column standardization (to standardize different data fields into a uniform format). Variables having an intrinsic ordered scale are encoded by ordinal, whereas nominal variables (e.g.,

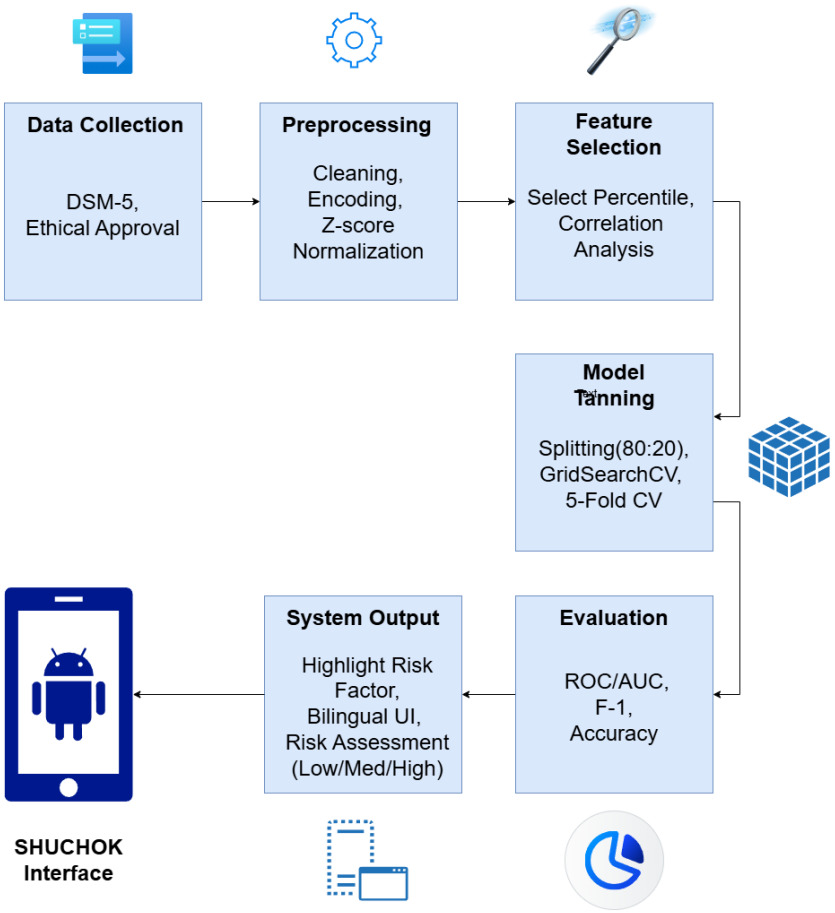

Figure 3.1: Conceptual Pipeline of the ADHD Prediction System Workflow

gender, parental status) are encoded using one-hot to prevent the artificial establishment of ordinal relationships. Encoding is done on categorical variables; that is, they are not numerical quantities but fixed groups. To ensure that the model is not misled by factors whose numerical values have larger values, numerical variables such as age and income are normalized to a similar range with z-score normalization. This transformation standardizes each feature to have a mean of zero and a standard deviation of one, ensuring equal contribution of all features to the model. The normalization is defined as:

$$z = \frac{x - \mu}{\sigma}$$

here x is the original feature value, μ is the mean, and σ is the standard deviation of the feature. This normalizing procedure makes all the features equally contributing to the learning process of the model possible [52]. Also, feature engineering that includes creating new and more informative features out of existing ones is another important aspect of this step. Examples are binary medical history flags and total behavioral scores (e.g., combining many items inattention). This comprehensive and systematic preparation process is important to improve data integrity, reduce possible biases, and to pre-process the

data to achieve optimal performance with different machine learning algorithms [53]. Machine learning models may give out noise and inconsistent results due to noise and the lack of consistent data incorrect or misleading results unless preprocessing is carefully performed [53].

The data proceeds to the feature selection step after thorough preprocessing. Among the enlarged set of features, this phase is vital in defining the most salient and predictive behavioral, demographic, and environmental features. Improving model interpretability, reducing computational cost, and reducing drastically the probability of overfitting are the primary goals of feature selection, particularly given the relatively small sample size of the study [23] [24] [25]. In the study, 13-20 predictors were found ideal as performance at varying percentiles (25, 50, and 75) was assessed. Besides socioeconomic factors such as money, they also incorporated such key features as age, gender, parental status and some behavioral indicators such as focus ability, distraction level, attention level, excessive talking, fidgeting, self-control, study focus, aggravation level and task omission. This rigorous feature selection process enhances predictive accuracy and applicability of models in Bangladeshi context by ensuring that models focus on most important data points [54].

Then, the re-filtered data set (that now contains the optimal features) is copied over to the Data Splitting phase where it is selectively split into an 80% training set and a 20% testing set. This split is very important to achieve the original class distribution and prevent biased learning of the models, and to ensure that the ratio of the ADHD and non-ADHD cases is always constant across both sets. The divided data then gets utilized in the Model Training stage. Four well-known machine learning algorithms - Random Forest, SVM, Logistic Regression and Decision Tree - are trained at this stage. The GridSearchCV method is applied with the five-fold cross-validation to tune hyperparameters so as to optimize the performance of each model and enhance the generalizability. To prevent overfitting and ensure that the model is only reliable when trying untested data, in this robust approach, a systematic search is made to acquire the best set of hyperparameters, by confirming the performance of the model on various subsets of the training data. To test the performance of the models on the test data that cannot be seen, the models undergo Model Evaluation to assess their performance. The most significant predictors are then identified using the Feature Relevance analysis which identified that the Random Forest model was the most important due to its high accuracy, F1-score, and AUC.

The final section of this system architecture is ADHD Prediction and conceptual User Interface. This interface is a prototyped interface in Canva (which will be both userfriendly and bilingual, i.e., both English and Bengali) that will act as an interface between end users such as parents, teachers, or guardians and the complicated machine learning backend. It aims to be a conveniently accessible awareness and self-assessment instrument that allows individuals to provide behavioral answers to the questionnaire

and get a rough idea of their ADHD risk level. Through a low-cost and culturally aware system to conduct a first ADHD screening, this interface will empower communities, which will eventually lead to an earlier diagnosis and allow timely treatment. Figure 3.1 depicts the entire system workflow of the system.

3.2 Dataset Description

3.2.1 Survey Design

The empirical basis of this study is done with systematic design and administration of a questionnaire which is the main instrument to get the data comprehensively. The questionnaire was designed in strict compliance with the diagnostic criteria of Diagnostic and Statistical Manual of Mental Disorders [5], given the importance of the necessity to adapt the questionnaire to the cultural requirements and relevance to the context. It is a universally agreed criterion that offers a sound empirically grounded framework on defining the main symptoms of ADHD, which ensures the clinical validity of the study in addition to facilitating the customization of the study to contextual requirements [1] [4] [6]. Notably, the DSM-5 tool has been fully translated into the Bangla language and adapted to the sociocultural context and language of the Bangladeshi population for its typical behavioral manifestations. This rigorous adaptation process enhanced the ecological validity and representativeness of the data collected significantly by ensuring that every question was both interpreted in the manner it had been designed to by the local participants and was also able to effectively measure the expression of behavioral characteristics in their specific cultural setting. The non-clinical, survey-based design of this instrument is a deliberate strategic choice that is aimed at circumventing the major barriers of expenses, limited access to special clinicians, and remote location, which often impede traditional diagnostic pathways in resource-limited settings [19] [20] [22]. This method enables the assessment tool to be more scalable and accessible to a larger population. Figure 3.2 gives a general layout of the survey questionnaire based on categories.

The survey per se was well broken down into five sections, each of which was supposed to draw in-depth, nonclinical information pertaining to potential indicators of ADHD. It was a multidimensional, comprehensive approach that was employed to collect as much information as possible on the functioning of the child, strategically going beyond the mere description of basic clinical manifestations to encompass a plethora of factors influencing the environment, development, and psychological aspects. Core Behavioral Symptoms were covered in the first segment. This section contained a set of well-worded questions, which overtly explored the behaviors that were indicative of impulsivity, hyperactivity, and inattention [1] [2] in line with the stipulated DSM-5 diagnostic criteria. Questions regarding the inability of a child to work or play in an organized way, observable cases

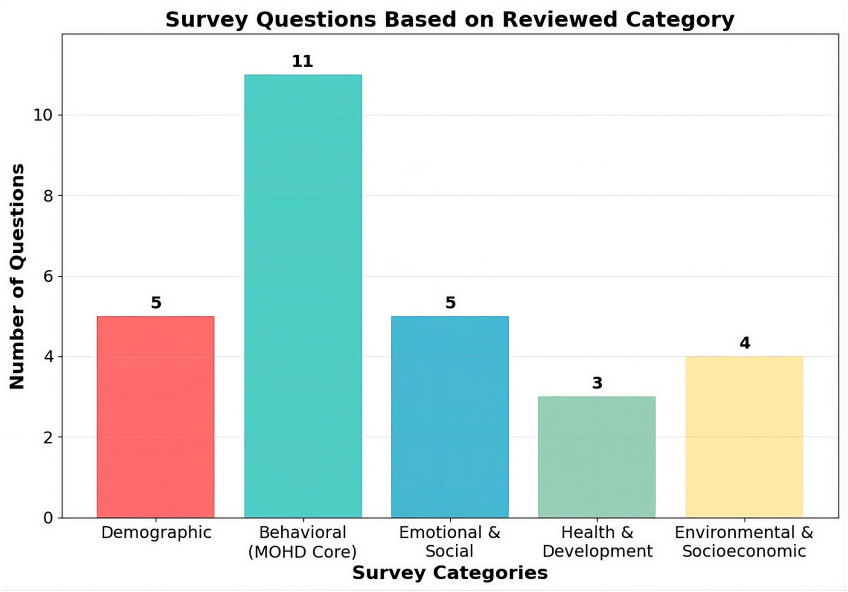

Figure 3.2: Survey Question Diagram

of fidgeting or wiggling of the seats, tendencies prejudice to excessive talking, and an inclination to cut off others were some of the particular instances. The questions of this section were vital in having a systematic finding as to whether these core ADHD traits exist or not, and also to what extent they exist. The second section of the survey focused on the systematic collection of key demographic information. This also contained some very vital information such as the age and gender of the child and both of these are known to be key epidemiological factors that affect the prevalence and manifestation of ADHD in different countries around the globe. It often demonstrated a great gender difference in which boys are diagnosed much more often than girls [3] [13]. In order to allow the exploration of more sociodemographic implications on ADHD presentation, other demographic details, including parental status (e.g., both biological parents were present in the household), were also meticulously recorded.

The third section involved Developmental and Health History. This part was designed to collect valuable information concerning critical developmental milestones in the early development of the child, such as walking, talking, and to identify any significant health-related events, chronic illnesses, or underlying medical conditions. Such history data is necessary in order to differentiate between ADHD symptoms and other developmental delays or health issues that might be similar to or aggravate the appearance of ADHD-like behaviors. The fourth segment was focused on the environmental and socioeconomic factors. This part of the questionnaire aimed to explore fully the immediate living condition of the child, socioeconomic status of the family, and the bigger background of the family. Questions in this section could have covered numerous areas, including family income levels, type of family structure (nuclear, extended, or single parent), and other potential environmental stressors or positive aspects in the life of the child. As existing

literature suggests, educational background and socioeconomic status of parents can play a major role in influencing the prevalence of ADHD and comorbid conditions [3] and this knowledge is invaluable when it comes to finding the predicted traits. Finally, Emotional or Social Difficulties were also discussed in the fifth and final section in a systematized manner. To get to know more about the child experiences the internalizing symptoms such as anxiety, externalizing symptoms such as aggression, problems with social contact and the general aggravation level, some questions were added. Since ADHD tends to be comorbid with other mental conditions and these emotional and social problems may severely affect the overall functioning of a child, academic performance, and psychosocial quality of life, these factors are highly significant [7] [31].

To primary informants such as parents, teachers and guardians, the survey was carefully designed, ensuring that each question was posed in friendly, non-clinical and culturally acceptable manner, that simplified the process of data collection. This is a very scalable and cost-effective approach to preliminary assessment of ADHD risk which is particularly useful in regions where professional diagnostic tools are limited.

3.2.2 Ethical Approval

Ethical integrity and responsible conduct of the research ensured the maximum protection of the rights, welfare and secrecy of all respondents. The appropriate Institutional Review Board (IRB, Independent University, Bangladesh) fully endorsed the entire study protocol, with the precise survey design, the meticulous data collection procedures and the intended data analysis methods. When all adult participants (including parents, guardians, and teachers) were fully informed about the study, strict procedures were established to ensure that no data was collected formally. Moreover, verbal consent was requested to older children where it was suitable and viable and their institution was acknowledged. This careful process made sure that all the participants were fully aware of the stated aim of the study, the concrete procedures to be conducted, any risks or discomforts that may arise and the benefits expected. It also made it clear beyond doubt that involvement was fully voluntary and was not subject to penalty at any time. In the entire research project, the privacy of the participants was safeguarded by the strict implementation of full secrecy and anonymity of all the obtained data.

3.2.3 Data Collection

This study uses 110 participants who were carefully selected and were between the ages of 4 and 17. The pool of participants was intentionally diverse and consisting of a variety of informants: parents, teachers, guardians and in case of developmental appropriateness, the older children themselves. This multi-informant approach is a foundation of rigorous ADHD assessment since it provides a comprehensive, ecological approach to the

behavioral profile of a child in multiple environments, such as the home, school, and social environment. This is vital in understanding the ubiquity of ADHD symptoms [3]. Despite its common usage in epidemiological studies, parental-reported status of ADHD has been known to possess a number of limitations, including not being able to share the method variation with other parent-reported predictor variables [45].

The case group of the study was a group of 44 children among the total subjects with a formal, pre-existing clinical diagnosis of ADHD. The other 66 kids (as a vital control group) were classified as non-ADHD. This balanced distribution made it possible to construct and do comprehensive testing of machine learning models that could reliably distinguish between these two vital categories. The subgroup of males with an ADHD diagnosis was 40 and included only 4 females, which was a notable feature of the collected data. This research is consistent with the established global epidemiological trends which have been consistently reporting higher diagnosis rates in boys. These patterns are often due to more overtly manifested, hyperactive-impulsive symptom presentation by boys that is easier to notice by adults [13]. Girls, however, tend to exhibit more insidious, undetective symptoms, which are misdiagnosed or overlooked, leading to an underdiagnosis [3]. Also, studies indicate that boys and girls might have different symptoms of ADHD during screening, which is why it is important to consider differences in the presentation of these conditions by gender [27]. Table 3.1 provides a summary of the demographics of the participants.

Table 3.1: Participant Summary Table

Characteristic	**ADHD**	**Non-ADHD**	**Overall**
Count	44	66	110
Mean Age	8.61 (3.24)	11.38 (2.44)	10.27 (3.09)
M/F	40/4	15/51	55/55

Additional demographic data analysis revealed that the two groups were different in terms of age. The mean age of the children with ADHD was 8.61 years (SD = 3.24 years). Conversely, non-ADHD control group was older with a mean age of 11.38 years (SD = 2.44 years). To obtain a rich, local, dataset to be used subsequently in the machine learning model development and validation, the highly attentive data collection process aimed to obtain responses that would be reflective of these nuanced behavioural and demographic features of the specific pediatric population in the Bangladeshi setting. Figure 3.3, 3.4, and 3.5 present the survey form that will be used in this study to systematically collect demographic and behavioral data of the participants.

Survey on ADHD Experiences and Early Recognition

Department of Computer Science & Engineering

Independent University, Bangladesh

Section 1: Details of informant / তথ্যদাতার বিস্তারিত

Are you filling out this survey on your own behalf? / আপনি কি এই প্রশ্নপত্র নিজের জন্য পূরণ করছেন?

☐ Yes/হ্যাঁ

☐ No/ না

If "No", please specify your relationship / যদি না হয়, দয়া করে সম্পর্ক উল্লেখ করুন:

☐ Parent / অভিভাবক

☐ Guardian / রক্ষক

☐ Teacher / শিক্ষক

☐ Care Giver

☐ Other:

Your Age Group / আপনার বয়সের পরিসর:*

☐ Under 25 / ২৫ এর নিচে

☐ 25–34 / ২৫–৩৪

☐ 35–44 / ৩৫–৪৪

☐ 45 and above / ৪৫ এবং তদূর্ধ্ব

☐ Other:

Your Education Level / আপনার শিক্ষাগত যোগ্যতা:*

☐ Primary / প্রাথমিক

☐ Secondary / মাধ্যমিক

☐ Higher Secondary / উচ্চ মাধ্যমিক

☐ Graduate / স্নাতক

☐ Postgraduate / স্নাতকোত্তর

☐ Other / অন্যান্য

_____

District/Area of Residence / বসবাসের জেলা/এলাকা:

_____

_____

Section 2: Child's Information / শিশুর তথ্য

Name of School / বিদ্যালয়ের নাম*:

_____

_____

Child's Age / শিশুর বয়স:

_____

Child's Gender / শিশুর লিঙ্গ:

☐ Male / ছেলে

☐ Female / মেয়ে

☐ Other / অন্যান্য

_____

Type of School / বিদ্যালয়ের ধরন:

☐ English Medium / ইংরেজি মাধ্যম

☐ Bengali Medium / বাংলা মাধ্যম

☐ Madrasa / মাদ্রাসা

☐ Homeschooling / বাসায় পাঠদান

☐ Other / অন্যান্য

_____

Number of siblings the child has / শিশুর ভাইবোনের সংখ্যা:

_____

Are both parents currently living with the child? / শিশুর সাথে উভয় পিতা-মাতা কি একসাথে বসবাস করেন?

☐ Yes / হ্যাঁ

☐ No / না

☐ Other / অন্যান্য

_____

Section 3: Behavioral Questions (DSM-5 Inspired) / আচরণভিত্তিক প্রশ্নাবলী (DSM-5 অনুপ্রাণিত)

Have you/your child often faced these issues in the last 6 months? / গত ৬ মাসে আপনি/আপনার সন্তান কি প্রায়ই এই সমস্যাগুলো সম্মুখীন হয়েছেন?

Section A: Inattention / মনোযোগের অভাব

Is it hard to maintain focus during class?
ক্লাসে মনোযোগ ধরে রাখতে কষ্ট হয় কি?
☐ Never (কখনই না)
☐ Sometimes (মাঝে মাঝে)
☐ Often (প্রায়ই)
☐ Always (সবসময়)

Is attention easily diverted to unrelated things?
সহজেই অন্য কোনো বিষয় মনোযোগ সরিয়ে নেয় কি?
☐ Never (কখনই না)
☐ Sometimes (মাঝে মাঝে)
☐ Often (প্রায়ই)
☐ Always (সবসময়)

Are tasks or homework often forgotten or abandoned midway?
টাস্ক বা হোমওয়ার্ক করতে গিয়ে মাঝপথে ভুলে যায় বা ছেড়ে দেয় কি?
☐ Never (কখনই না)
☐ Sometimes (মাঝে মাঝে)
☐ Often (প্রায়ই)
☐ Always (সবসময়)

Figure 3.3: Survey Form

Is it difficult to understand or pay attention to what is said in class?
ক্লাসে যা বলা হচ্ছে তা বুঝতে বা খেয়াল রাখতে কষ্ট হয় কি?
☐ Never (কখনই না)
☐ Sometimes (মাঝে মাঝে)
☐ Often (প্রায়ই)
☐ Always (সবসময়)

Is it hard to pay attention to details?
বিস্তারিত বিষয়গুলিতে মনোযোগ দেওয়া কঠিন হয় কি?
☐ Never (কখনই না)
☐ Sometimes (মাঝে মাঝে)
☐ Often (প্রায়ই)
☐ Always (সবসময়)

Is it often difficult to keep track of things needed for tasks or activities?
কাজ বা ক্রিয়াকলাপের জন্য প্রয়োজনীয় জিনিসগুলি প্রায়ই হারিয়ে যায় কি?
☐ Never (কখনই না)
☐ Sometimes (মাঝে মাঝে)
☐ Often (প্রায়ই)
☐ Always (সবসময়)

Section B: Hyperactivity & Impulsivity / অতিসক্রিয়তা ও আকস্মিক আচরণ

Is it difficult to sit still for a long time?
দীর্ঘ সময় চুপচাপ বসে থাকা কঠিন হয় কি?
☐ Never (কখনই না)
☐ Sometimes (মাঝে মাঝে)
☐ Often (প্রায়ই)
☐ Always (সবসময়)

Are things often blurted out unnecessarily?
অপ্রয়োজনে কথা বের হয়ে যায় কি?
☐ Never (কখনই না)
☐ Sometimes (মাঝে মাঝে)
☐ Often (প্রায়ই)
☐ Always (সবসময়)

Is waiting for a turn difficult?
ক্রম বা অপেক্ষা করতে কষ্ট হয় কি?
☐ Never (কখনই না)
☐ Sometimes (মাঝে মাঝে)
☐ Often (প্রায়ই)
☐ Always (সবসময়)

Are actions often done without thinking?
বিনা চিন্তায় কাজ করা হয় কি?
☐ Never (কখনই না)
☐ Sometimes (মাঝে মাঝে)
☐ Often (প্রায়ই)
☐ Always (সবসময়)

Is there easy irritation or restlessness?
সহজেই বিরক্তি বা অস্থিরতা হয় কি?
☐ Never (কখনই না)
☐ Sometimes (মাঝে মাঝে)
☐ Often (প্রায়ই)
☐ Always (সবসময়)

Is there frequent fidgeting or movement when sitting still is expected?
চুপচাপ বসে থাকার সময় প্রায়ই অস্থিরতা বা নড়াচড়া হয় কি?
☐ Never (কখনই না)
☐ Sometimes (মাঝে মাঝে)
☐ Often (প্রায়ই)
☐ Always (সবসময়)

Is there talking more than usual or excessive talking?
সাধারণের চেয়ে বেশি বা অতিরিক্ত কথা বলা হয় কি?
☐ Never (কখনই না)
☐ Sometimes (মাঝে মাঝে)
☐ Often (প্রায়ই)
☐ Always (সবসময়)

Section C: Social & Emotional / সামাজিক ও আবেগিক দিক

Is there difficulty socializing or feelings of loneliness?
বন্ধুদের সাথে সহজে মিশতে সমস্যা হয় বা একাকীত্ব লাগে কি?
☐ Never (কখনই না)
☐ Sometimes (মাঝে মাঝে)
☐ Often (প্রায়ই)
☐ Always (সবসময়)

Is there frequent loss of self-control?
নিজেকে নিয়ন্ত্রণ করা কঠিন হয় কি?
☐ Never (কখনই না)
☐ Sometimes (মাঝে মাঝে)
☐ Often (প্রায়ই)
☐ Always (সবসময়)

Are there signs of aggression or emotional outbursts?
আগ্রাসন বা আবেগপ্রবণ আচরণের লক্ষণ দেখা গেছে কি?
☐ Never (কখনই না)
☐ Sometimes (মাঝে মাঝে)
☐ Often (প্রায়ই)
☐ Always (সবসময়)

Is there frequent anxiety or worry?
অনেক সময় চিন্তিত বা দুশ্চিন্তা হয় কি?
☐ Never (কখনই না)
☐ Sometimes (মাঝে মাঝে)
☐ Often (প্রায়ই)
☐ Always (সবসময়)

Section D: Additional / অতিরিক্ত

Is studying affected due to lack of focus?
পড়াশোনায় মনোযোগ কম হওয়ার কারণে সমস্যা হয় কি?
☐ Yes (হ্যাঁ)
☐ No (না)
☐ Not sure (নিশ্চিত নই)

Has there been any prior medical evaluation for ADHD?
আগে কখনো ADHD নিয়ে ডাক্তার দেখানো হয়েছে কি?
☐ Yes (হ্যাঁ)
☐ No (না)
☐ Don't remember (মনে নেই)

Figure 3.4: Survey Form

Section 4: Environmental and Lifestyle Factors / পরিবেশ ও জীবনযাত্রা

Does the child have any known health conditions (e.g., asthma, epilepsy, allergies, etc.)? / শিশুর কি কোনো বিদ্যমান শারীরিক সমস্যা আছে (যেমন: হাঁপানি, মৃগী, অ্যালার্জি ইত্যাদি)?

_____

Is there any family history of genetic or neurological disorders (e.g., ADHD, autism, learning disability, mental illness)? / আপনার পরিবারের কারো মধ্যে কি বংশগত বা স্নায়বিক রোগ (যেমন: ADHD, অটিজম, শিখনে সমস্যা, মানসিক সমস্যা) আছে?

_____

How many hours does the child spend on screen per day? / প্রতিদিন শিশু কত ঘণ্টা স্ক্রিনের (টিভি, মোবাইল, কম্পিউটার) সামনে থাকে?

☐ Less than 1 hour / ১ ঘণ্টার কম

☐ 1–2 hours / ১–২ ঘণ্টা

☐ 3–4 hours / ৩–৪ ঘণ্টা

☐ 5–6 hours / ৫–৬ ঘণ্টা

☐ 6+ hours / ৬ ঘণ্টার বেশি

☐ Other:

How many hours does the child sleep at night on average? / গড়ে রাতে শিশু কত ঘণ্টা ঘুমায়?:

_____

Does the child have access to physical activity/play daily? / শিশু কি প্রতিদিন শারীরিক কার্যকলাপ বা খেলার সুযোগ পায়?

☐ Yes / হ্যাঁ

☐ No / না

Has the child been exposed to major family stress (e.g., divorce, death, financial problems)? / শিশু কি কোন বড় পারিবারিক চাপের সম্মুখীন হয়েছে (যেমন: বিবাহ বিচ্ছেদ, মৃত্যু, আর্থিক সমস্যা)?

☐ Yes / হ্যাঁ

☐ No / না

Does the child live in a joint family or nuclear family? / শিশু কি যৌথ পরিবারে না কি একক পরিবারে বসবাস করে?

☐ Joint / যৌথ

☐ Nuclear / একক

Section 5: Socioeconomic Information / আর্থসামাজিক তথ্য

What is your monthly family income range? / আপনার মাসিক পারিবারিক আয় কত?

☐ Less than 15,000 BDT / ১৫,০০০ টাকার কম

☐ 15,000 – 30,000 BDT / ১৫,০০০ – ৩০,০০০ টাকা

☐ 30,001 – 50,000 BDT / ৩০,০০১ – ৫০,০০০ টাকা

☐ 50,001 – 75,000 BDT / ৫০,০০১ – ৭৫,০০০ টাকা

☐ Above 75,000 BDT / ৭৫,০০০ টাকার বেশি

Highest level of education among parents / শিশুর পিতা-মাতার মধ্যে সর্বোচ্চ শিক্ষাগত যোগ্যতা:

☐ No formal education / আনুষ্ঠানিক শিক্ষা নেই

☐ Primary / প্রাথমিক

☐ Secondary / মাধ্যমিক

☐ Higher Secondary / উচ্চ মাধ্যমিক

☐ Graduate/Postgraduate / স্নাতক/স্নাতকোত্তর

Figure 3.5: Survey Form

3.2.4 Tools and Technologies

The entire project was developed in Python [55], which is highly adaptive and popular programming language, for data cleaning, complex data preprocessing, proper feature selection, and to completely develop and test the machine learning models. All computations for hyperparameter optimization and cross-validation were done using Google Colaboratory [56], a Jupyter notebook service hosted in the cloud that allowed us to access computing resources without overloading our local machines.

The ability to manage, structure and analyse data efficiently and in depth made **pandas** (v2.2) [57,58] the key library for this technical architecture as it enabled the easy processing of tabular survey data, such as loading, missing value imputation, encoding of categorical variables, and generation of descriptive statistics. With the introduction of **NumPy** (v2.0) [59] the numerical operations were made possible with the help of efficient numerical tools, especially the multi-dimensional arrays which are the base for all the machine learning calculations carried out in this study.

As an auxiliary tool for feature selection, exploratory regression diagnostics and for statistical modelling, **SciPy** (v1.14) [60] and **statsmodels** (v0.14) [61] were used.

For visualisation of the experimental results, **Matplotlib** (v3.10) [62] was used to generate the ROC curves, bar charts, and feature-importance plots shown in Chapter 5, while **Seaborn** (v0.13) [63] was used for the correlation heatmaps and the colour-coded confusion-matrix grids.

Finally, model persistence was handled by **joblib** (v1.4) [64] and Python's built-in `pickle` module, which serialised the trained classifiers and label encoders to disk so that the same models could be reloaded without retraining for use in the proposed mobile application (Chapter 4).

This combination of mature, open-source scientific computing libraries reflects an iterative development technique that was resource-intensive in the beginning but eventually enhanced the efficiency of the computing and predictive accuracy of the selected machine learning algorithms [65].

3.3 Data Preprocessing and Feature Selection

To ensure the data quality, consistency, and applicability to machine learning algorithms, the obtained raw data in the form of the culturally adapted survey went through a vigorous and systematic data preprocessing process. This is a vital step since it transforms raw, potentially noisy, incomplete and inconsistently formatted responses to a clean, structured and analysable dataset to significantly affect the accuracy, dependability and interpretability of the resulting prediction models [53] [66]. Without proper preprocessing, machine learning algorithms might have a hard time identifying meaningful patterns that

lead to the wrong conclusion or not optimal performance [67]. Preprocessing pipeline steps such as data cleansing, encoding categorical variables, scaling numerical features and derived features development were all significant steps in the pipeline.

Data cleaning was initially done in order to rectify inconsistencies and ensure consistency. The primary focus of this was whitespace normalization of all textual and categorical entries. As an illustration, varying placement of letters or varying number of spaces might lead to varying interpretation of different algorithms, although the meaning underlying the same algorithm might have been the same. Resolving these small differences was essential to minimize data distortion and prevent similar items from being mis-classified. To further enhance the quality of the data, the procedure also indirectly involves correcting any glaring errors or errors in entries. To encode continuous data into a numerical format that could be processed by machine learning algorithms, categorical variables needed to be encoded. Following the intrinsic nature of the dataset, 27 categorical features of the dataset were processed with the help of two encoding strategies:

- Nominal categorical variables without an innate ordinal relationship were subjected to one-hot encoding. In each unique category within a feature, this approach results in the creation of new binary columns, with a 1 representing the appearance of a category and a 0 representing the absence of a category. This retains categories that are fundamentally distinct with giving the algorithm an artificial sense of order or scale [68].

- Ordinal encoding was used on a limited scale to describe categorical variables which intuitively have an inherent order or rank. Assigning each category a unique number, this approach preserves the significant order and allows the algorithm to capitalize on this natural hierarchy in the process of learning [68]. As an illustration, the 27 categorical variables were label encoded and the target variable was label encoded as a binary outcome (ADHD = 1, non-ADHD = 0).

Numerical features were encoded and then scaled using z-score normalization. Examples of numerical characteristics that may have very different scales, and units include age, income, aggregate behavioral scores. This difference can lead to some machine learning algorithms which are sensitive to the magnitude of features to prefer features with more numerical values during the learning process. Z-score normalization is used to convert features to have a mean of 0 and a standard deviation of 1. This process ensures that every numerical feature can contribute equally to the weights and gradients of the model, which improves the convergence and the overall performance by ensuring that none of the features influences the model unduly [67] [69]. Finally, to enhance the prediction and insight generation capabilities of the dataset, derived features, or feature engineering, were developed. To include more complex patterns or correlations that may not be readily

apparent, this includes deriving new, more useful features by including existing raw or processed features. Binary medical history flags and aggregate behavioral ratings are two examples. This intentional generation of higher-level features can significantly improve the performance of models since the algorithms can then act upon more meaningful representations of the underlying phenomena [70]. This process yielded the final target vector (y axis) and the final feature matrix (x axis), denoting the cleaned, processed, and fully-prepared set of data that would be fed into the model for the next stages of model training and model evaluation.

Recursive Feature Elimination (RFE) and **SelectPercentile** were used to identify the most significant indicators (attention level, fidgeting, study focus, etc.).

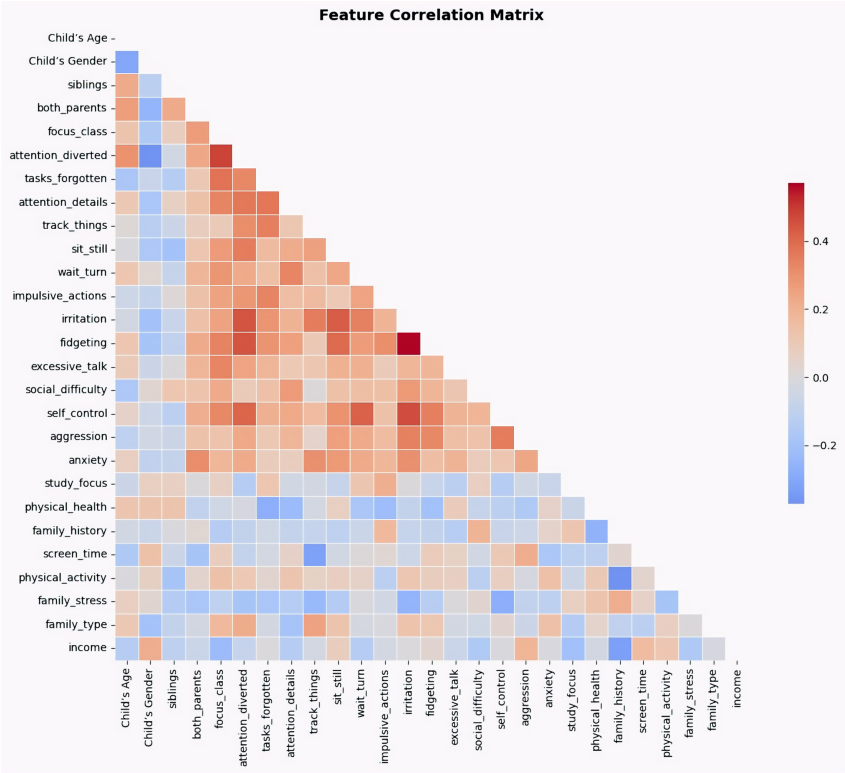

Figure 3.6: Feature Correlation Matrix

The aim of feature selection which is a crucial step in the machine learning process is to find the characteristics that are the most relevant and most important, out of the already processed dataset. The main objective of this procedure is to achieve some important objectives: to reduce the computation cost of prediction and training, to make the model more interpretable by simplifying the input space, and, above all, to minimize the risk of overfitting, which is particularly important bearing in mind the moderate sample size of the study [23] [24] [25]. By focusing on a set of the best predictors, the models also have the ability to learn more robust and generalized patterns that enhance the accuracy of predictions on unseen data [71]. This was achieved by applying a multifaceted approach that involved correlation analysis, SelectPercentile, and Recursive Feature Elimination to

the project.

- Recursive Feature Elimination: This method step-by-step identifies the most relevant features by repeatedly training a machine learning model and removing the least important features at each stage. Each stage of process that begins with all features is evolved and feature importances are computed. After the feature least relevant is eliminated, the model is retrained with the smaller feature set. This cycle continues until the optimal number of features - determined by cross-validation - is achieved [72].

Table 3.2: Feature Selection Results by SelectPercentile

Metric	**Value**	**Details / Selected Features**
Percentile: 25%		
Number of Features	7	
Selected Features		Child's Age, Child's Gender, Parental status, Focus ability, Distraction level, Attention level, Excessive talking
Mean Accuracy	N/A	
Percentile: 50%		
Number of Features	13	
Selected Features		Child's Age, Child's Gender, Parental status, Focus ability, Distraction level, Tasks omission, Attention level, Irritation level, Fidgeting level, Excessive talking, Self control, Study focus, Income
Mean Accuracy	N/A	
Percentile: 75%		
Number of Features	20	
Selected Features		Income, Siblings, Sit_still, Anxiety, Family_type, Track_things, Impulsive_actions, Physical_activity, Physical_health, Aggression, Screen_time, Wait_turn, Family_history, Social_difficulty, Family_stress, Study_focus, Irritation, Self_control, Tasks_forgotten, Fidgeting
Mean Accuracy	0.8727 ± 0.0872	

Feature Selection Results by SelectPercentile

- Correlation Analysis: This method found features that had significant statistical correlations with the target variable. The study was able to get preliminary, understandable results on possible predictors calculating the Pearson correlation coef-

ficients between each candidate trait and the result of ADHD. Features with higher value of absolute correlation were discussed as the ones that provided a stronger linear relationship with the absence of ADHD, and their strength was considered stronger. This study has identified several factors that are significant as we can see in Figure 3.6.

- SelectPercentile: This approach, based on a univariate test statistic, gives a customized option of picking a specified percentage of the highest rated characteristics. It rates each characteristic after the individual ranking of the characteristic in relation to the target variable. This allows the user to specify a percentile in order to retain the best features. This approach will help in the regulated dimensionality reduction. In this research, SelectPercentile was evaluated at 25, 50 and 75 percentages. In Table 3.2, it is demonstrated that selecting a set of 20 features (at the 75 percentile) had the highest scores and the average cross-validation accuracy of 5-folds measures 0.8727 with a STANDARD DEVIATION of 0.0872. This was effective in finding a compromise between its predictive power and reducing the dataset dimensionality bt 25.9%.

Final 20 features to include in model training were income, siblings, sitstill, anxiety, family type, trackthings, impulsive actions, physical activity, physical health, aggression, screen time, wait turn, family history, social difficulty, family stressed, study focus, irritation, self control, forgot things, and fidgeting (at the 75 percentile level, as represented in Table 3.2. These attributes were determined as most significant with regards to predicting ADHD in the Bangladesh context, in addition to Child's Gender, attention diverted, fidgeting, study focus and irritation (identified as the strongest predictors of ADHD by correlation analysis). This made sure that the machine learning models obtained focused on the most important and contextually meaningful data points [40].

Mathematical Formulation of SelectPercentile: Feature selection in this study was carried out using the *SelectPercentile* method with the ANOVA F-statistic (`f_classif`) as the scoring function. For each feature X_j ($j = 1, 2, \ldots, p$), the F-statistic measures how strongly the feature varies across the two target classes (ADHD vs. non-ADHD), defined as the ratio of between-group variance to within-group variance:

$$F_j \;=\; \frac{\mathrm{MS}_{between}}{\mathrm{MS}_{within}}, \tag{3.1}$$

where

$$\mathrm{MS}_{between} \;=\; \frac{1}{k-1}\sum_{c=1}^{k} n_c\,(\bar{x}_{j,c} - \bar{x}_j)^2, \tag{3.2}$$

$$\mathrm{MS}_{within} \;=\; \frac{1}{n-k}\sum_{c=1}^{k}\sum_{i \in c}(x_{ij} - \bar{x}_{j,c})^2. \tag{3.3}$$

Here k indicates the number of classes ($k = 2$ in our case), n_c is the number of samples in class c, $\bar{x}_{j,c}$ is the sample mean of feature j within class c, and $\bar{x}_j$ is the overall mean of feature j across all n observations. A higher F-value indicates that the feature contributes greater discriminative power between the ADHD and non-ADHD groups, since the between-class variation dominates the within-class variation.

SelectPercentile ranks all features by their F-statistic F_j and retains only the top $P\%$. In this study three percentile settings were evaluated, $P \in \{25, 50, 75\}$, as reported in Table 3.2; the configuration that retained the top 75% of features ($p' = 20$) yielded the highest mean cross-validation accuracy (0.8727 ± 0.0872) and was therefore adopted for the subsequent model-training stage.

3.4 Data Splitting

The ready dataset was systematically divided into discrete subsets to allow building machine learning models and to objectively evaluate them. The procedure included splitting the data into 80 and 20 percent, respectively. In machine learning, this popular partition ratio is used to make sure that much of the data is accessible to the models to learn more complex patterns and relationships and leaves a separate, unseen part of the data to be able to evaluate their capability to extrapolate to new and unobservable cases [25]. In order to prevent an exaggeration of the effectiveness of the model against the actual world, as shown in Figure 3.7 the test set is used as a objective measure of the performance of the model of the data that it was not exposed to during training [45].

5 FOLDS CV PROCESS FLOW					
Validation (V)/Tr-anning (T)	1	2	3	4	5
Iteration 1	V	T	T	T	T
Iteration 2	T	V	T	T	T
Iteration 3	T	T	V	T	T
Iteration 4	T	T	T	V	T
Iteration 5	T	T	T	T	V

Figure 3.7: 5-fold cross-validation

Stratification was also a significant and substantially considered aspect of our data separation method. Stratification ensures that the distribution of the training and testing sets is also represented in the original total dataset by ensuring that the proportion of

instances is constant across each of the classes in the training and testing set [73] [74]. The preservation of this distribution of the class between splits is particularly important regarding classification tasks since the dataset presented was formed by 44 ADHD cases and 66 non-ADHD cases, which resulted in an unbalanced distribution of classes. In the absence of stratification, there is a possibility that a simple random division will accidentally result in test sets that contain either fewer or no representatives of the minority group. This may lead to biased model evaluation, since the model may work well on the majority but poorly on the minority and the measures that will be reported would fail to represent this weakness in an adequate manner [29]. Stratification ensures the trained models are presented with a representative sample of both classes during training and that the trained models are tested on a test set that sufficiently represents the real-world prevalence and challenges of making the distinction between these classes by preserving the original distribution of classes in both the ADHD and non-ADHD groups in the training and testing sets. This stringent data splitting and stratification forms the backbone of the development of machine learning models that are accurate, reliable, and applicable to new, unobserved data on the target population.

Mathematical Formulation of Stratified K-Fold CV: Because of the moderate size of the dataset ($n = 110$), Stratified K-Fold Cross-Validation with $K = 5$ was used to obtain robust performance estimates. Stratification is a technique that maintains the original class proportion (60% non-ADHD : 40% ADHD) within each fold. Formally, the dataset $\mathcal{D}$ is partitioned into K disjoint folds $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_K$ such that

$$\mathcal{D} \;=\; \mathcal{D}_1 \cup \mathcal{D}_2 \cup \ldots \cup \mathcal{D}_K, \qquad \mathcal{D}_i \cap \mathcal{D}_j \;=\; \emptyset \quad for i \neq j. \tag{3.4}$$

At iteration k, fold $\mathcal{D}_k$ is used as the test set and the remaining $K - 1$ folds form the training set. The cross-validated estimate of any metric M is then

$$\bar{M} \;=\; \frac{1}{K} \sum_{k=1}^{K} M_k, \tag{3.5}$$

with corresponding standard deviation

$$\sigma_M \;=\; \sqrt{\frac{1}{K-1} \sum_{k=1}^{K} (M_k - \bar{M})^2}\,. \tag{3.6}$$

Here M_k denotes the value of the chosen metric (Accuracy, Precision, Recall, Specificity, F1-Score, or AUC) computed on fold $\mathcal{D}_k$. The last reported performance of each model is expressed as $\bar{M} \pm \sigma_M$ A single train–test split would not provide as accurate an estimate through the five folds as would be provided by the number of times the data was split. This reduces the variation due to the small sample size.

3.5 ML Model Training

Training appropriate machine learning models is a critical step to developing a dependable system of predicting ADHD. In this step, a diverse set of algorithms were used to explore the underlying patterns and relationships that differentiate between ADHD and non-ADHD cases in Bangladeshi children, based on the thoughtfully preprocessed data and the thoughtfully chosen features. The primary aim was to find a model which was able to predict ADHD risk with good reliability based on the non-clinical survey-based data. To ensure sufficient evaluation and to leverage the benefits of various algorithmic paradigms, four machine learning architectures, such as Random Forest, Support Vector Machine, Logistic Regression, and Decision Tree, have been selected to train with them. The choice of these models was made in regards to how applicable they were to data of both this size and the type, how interpretable they were and how effective they were demonstrated to be in classification.

Let the dataset be denoted by $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$, where $\mathbf{x}_i \in R^p$ represents the feature vector of the i-th child (with $p = 27$ features after encoding) and $y_i \in \{0, 1\}$ is the binary class label, with $y_i = 1$ indicating ADHD and $y_i = 0$ indicating non-ADHD. In our study, $n = 110$.

Random Forest (RF): RF a powerful ensemble-learning method based on bagging and decision trees. In training, RF trains as many decision trees as configured. It has an inherent resistant to overfitting, especially in comparison with single decision trees due to its combination of different models. RF does not involve much preparation like feature scaling to work with high-dimensional and heterogeneous data, a combination of numerical and categorical data. Also, it provides a reliable measure of feature relevancy, showing the variables that have the most impacts on the predictions. RF models have persistently run more accurate than other models in applied forests with regards to predicting ADHD, and they are often able to achieve high accuracy and AUC [38] [40]. It is particularly suitable for behavioural and demographic data, where interactions between features can be subtle and multidimensional. Furthermore, it is capable of modelling complex non-linear relationships within the data.

Mathematical Formulation of RF: The (RF) is an ensemble learning technique that combines multiple Decision Trees to improve classification performance. An ensemble is formed by aggregating the predictions of B individual Decision Trees, each trained on a different subset of the training data. For each tree, a bootstrap sample is generated by randomly selecting data points from the original dataset with replacement. In addition, a random subset of features is considered at each split during the tree construction process. The combination of bootstrap aggregation (*bagging*) and random feature selection reduces the variance of individual trees and helps prevent overfitting, particularly when working with small or moderately sized datasets, such as the one used in this study.

Let $h_1(\mathbf{x}), h_2(\mathbf{x}), \ldots, h_B(\mathbf{x})$ denote the B individual decision trees trained in the ensemble (in our experiments $B = 100$, the scikit-learn default). For a binary classification problem, the predicted class of the Random Forest is Computing the majority vote over all the trees:

$$\hat{y}_{\text{RF}} \;=\; *mode_{b\in\{1,\ldots,B\}}\{\, h_b(\mathbf{x})\,\}. \tag{3.7}$$

Equivalently, the probability of class c estimated by the forest is the average of the per-tree class probabilities:

$$P(y = c \mid \mathbf{x}) \;=\; \frac{1}{B}\sum_{b=1}^{B} P_b(y = c \mid \mathbf{x}), \tag{3.8}$$

where $P_b(y = c \mid \mathbf{x})$ is the class-c probability returned by the b-th tree (computed as the proportion of class-c training samples in the leaf to which $\mathbf{x}$ falls, as defined in Equation 3.24.

Each individual tree h_b is constructed independently using two sources of randomness:

1. **Bootstrap sampling.** Each tree is trained on a bootstrap sample $\mathcal{D}_b$ of size n drawn with replacement from the original training set $\mathcal{D}$. Approximately 63.2% of the unique training samples appear in any given $\mathcal{D}_b$, while the remaining samples (the *out-of-bag* samples) can be used as an internal validation set.
2. **Random feature subset.** At each node of every tree, only a random subset of m features (with $m \approx \sqrt{p}$ for classification) is considered when searching for the optimal split. This decorrelates the individual trees and prevents a few dominant features from driving every tree to make similar predictions.

Within each tree, splits are chosen by minimising the Gini Impurity defined in Equation eq:dt-gini.

The theoretical foundations for the ensemble are based on the bound on the generalization error, obtained by Breiman [75]:

$$\text{PE}^* \;\leq\; \frac{\bar{\rho}\,(1 - s^2)}{s^2}, \tag{3.9}$$

where PE^* is the generalisation error of the forest, s is the *strength* of the individual classifiers (their expected margin on a random test sample), and $\bar{\rho}$ is the mean correlation between trees. Equation 3.9 shows that the upper bound on the forest's error is reduced when the individual trees are kept strong (large s) but de-correlated (small $\bar{\rho}$). These two sources of randomness outlined above apply these as practical ways that RF can enforce them trade-off, so the algorithm generally obtains lower rate of success.

Support Vector Machine (SVM): SVMs are powerful and versatile classification algorithms that are particularly effective in case where a discernible gap between the classes. The basic idea of SVM is to identify an optimal hyperplane in a high-dimensional

feature space that would separate the different classes in a maximum result. This maximum margin strategy enhances the generalization of the model. SVMs can solve both linear and non-linear classification problems using a combination of the different kernel functions (polynomial, radial basis functions, etc.), which implicitly project input properties into higher-dimensional spaces. Due to their adaptability, SVMs can recognize complex trends in the data that are not linear to decompose. SVMs have been widely applied to medical classification problems, including ADHD prediction. With appropriate calibration, they often achieve competitive accuracy and perform effectively even when trained on moderately sized datasets [35] [40]. Their ability to work on high-dimensional spaces and the ability of operating features of varying application sizes were the most important considerations when it comes to the selection of them.

Mathematical Formulation of SVM:

The Support Vector Machine (SVM) constructs an optimal separating hyperplane that maximises the geometric margin between the two classes. Because not all real-world datasets are linearly separable, the soft-margin formulation [76] allows a controlled amount of margin violation through non-negative slack variables ξ_i. For binary labels remapped internally to $y_i \in \{-1, +1\}$ (from the original $\{0, 1\}$ encoding), the primal training problem is the constrained convex optimisation

$$\min_{\mathbf{w}, b, \xi} \ \frac{1}{2}\,\mathbf{w}_2^2 \ + \ C \sum_{i=1}^{n} \xi_i, \tag{3.10}$$

subject to

$$y_i(\,\mathbf{w}^\top \phi(\mathbf{x}_i) + b\,) \ \geq \ 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, 2, \ldots, n. \tag{3.11}$$

Here $\mathbf{w}$ is the normal vector of the separating hyperplane, b is the bias, $\phi(\cdot)$ is a (possibly non-linear) feature map, and $C > 0$ is a regularisation parameter that controls the trade-off between margin maximisation and tolerance to misclassification: a larger C penalises violations more heavily and produces a tighter margin.

Direct optimisation of Equation 3.10 in the (often high-dimensional) feature space induced by $\phi(\cdot)$ would be computationally prohibitive. The problem is reformulated using Lagrangian duality, leading to the dual form, which is the form actually solved in practice:

$$\max_{\alpha} \ \sum_{i=1}^{n} \alpha_i \ - \ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \, \alpha_j \, y_i \, y_j \, K(\mathbf{x}_i, \mathbf{x}_j), \tag{3.12}$$

subject to $0 \leq \alpha_i \leq C$ for $i = 1, \ldots, n$ and $\sum_{i=1}^{n} \alpha_i \, y_i = 0$. The dual depends on the input samples only through the inner products $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j)$, which is the well-known *kernel trick*: the mapping $\phi(\cdot)$ never has to be evaluated explicitly.

For this study, the SVM was configured with the Radial Basis Function (RBF) kernel,

defined as:

$$K(\mathbf{x}_i, \mathbf{x}_j) \;=\; \exp\Big(-\gamma\, \mathbf{x}_i - \mathbf{x}_j{}_2^2\Big), \tag{3.13}$$

where $\gamma > 0$ controls the influence of a single training sample. A small γ produces a smooth, low-curvature decision boundary, whereas a large γ produces a more localised boundary that fits the training points tightly. Scikit-learn defaults to the heuristic $\gamma = 1/(p \cdot \mathrm{Var}(\mathbf{X}))$, which was retained in our experiments.

Once the dual variables α have been estimated, the decision function for a new sample $\mathbf{x}$ is

$$f(\mathbf{x}) \;=\; sign\left(\sum_{i=1}^{n} \alpha_i\, y_i\, K(\mathbf{x}_i, \mathbf{x}) + b\right). \tag{3.14}$$

Only training samples for which $\alpha_i > 0$ — the *support vectors* — contribute to the final classifier; all other samples can be discarded without affecting the decision boundary. This sparsity property is one of the main reasons SVMs generalise well even on small datasets.

Standard SVMs return only signed decision values $f(\mathbf{x})$, not probabilities. Because our evaluation relies on the AUC and ROC curves, which require probabilistic outputs, we used Platt scaling [77] to convert decision values into class probabilities by fitting a sigmoid to the SVM output:

$$P(y = 1 \mid \mathbf{x}) \;=\; \frac{1}{1 + \exp(\, A \cdot f(\mathbf{x}) + B\,)}, \tag{3.15}$$

where the parameters A and B are estimated by maximum likelihood on the training set (in scikit-learn this is enabled with `probability=True`). The SVM was selected for inclusion in the comparison because of its strong theoretical foundation in margin maximisation, its robustness in high-dimensional feature spaces, and its competitive performance on moderate-sized clinical datasets reported in the related ADHD-prediction literature [35] [40].

Logistic Regression (LR): LR remains a basic and extremely practical model to use in binary classification tasks, even though it is a relatively simplistic statistical model unlike RF or SVM. LR is a method that, based in a logistic function, attempts to determine the probability of an outcome (i.e., whether ADHD is present or not). Its key advantage is its interpretability; the coefficients of each feature provide a clear vision of the direction and the strength of their influence on the likelihood of the outcome. This transparency is necessary in clinical applications, in which it can be equally important to know the reason why a forecast is generated, rather than the prediction itself.

Mathematical Formulation of LR: LR is a discriminative linear classifier which models the conditional probability of the positive class using the sigmoid (logistic) function. Given a feature vector $\mathbf{x} = (x_1, x_2, \ldots, x_p)^\top$ with weight vector $\mathbf{w} = (w_1, w_2, \ldots, w_p)^\top$

and bias term b, the predicted probability that $\mathbf{x}$ belongs to the positive class (ADHD = 1) is

$$P(y = 1 \mid \mathbf{x}) \;=\; \sigma(z) \;=\; \frac{1}{1 + e^{-z}}, \tag{3.16}$$

where the linear combination z is given by

$$z \;=\; \mathbf{w}^{\top}\mathbf{x} + b \;=\; \sum_{j=1}^{p} w_j \, x_j + b. \tag{3.17}$$

The sigmoid function $\sigma(\cdot)$ maps the linear output z to the interval $(0, 1)$, allowing the result to be interpreted as a probability. The decision rule is

$$\hat{y} \;=\; \{\, 1 \,, if P(y = 1 \mid \mathbf{x}) \;\geq\; 0.5, 0, otherwise. \tag{3.18}$$

The model parameters $\mathbf{w}$ and b are estimated by minimizing the binary cross-entropy (log-loss) function on the training set:

$$\mathcal{L}(\mathbf{w}, b) \;=\; -\frac{1}{n} \sum_{i=1}^{n} \Big[\, y_i \, \log(p_i) \;+\; (1 - y_i) \, \log(1 - p_i) \,\Big], \tag{3.19}$$

where $p_i = \sigma(\mathbf{w}^{\top}\mathbf{x}_i + b)$ is the predicted probability for the i-th training sample. Scikit-learn applies an L_2 penalty by default, so the regularised objective that is minimised in practice is

$$\mathcal{J}(\mathbf{w}, b) \;=\; \mathcal{L}(\mathbf{w}, b) \;+\; \frac{1}{2C} \, \mathbf{w}_2^2, \tag{3.20}$$

where $C > 0$ is the inverse regularisation strength (a smaller C applies stronger regularisation) and $\mathbf{w}_2$ denotes the Euclidean norm of the weight vector. Optimisation is performed using the L-BFGS solver with a maximum of 1000 iterations. LR was selected as a baseline classifier in this study because of its strong interpretability: each coefficient w_j directly indicates the magnitude and direction of the contribution of feature X_j to the log-odds of an ADHD diagnosis.

Decision Tree (DT): DTs are basic rule-based models, which are based on a recursive algorithm, partitioning the data space into a collection of regions, each classfied with a class label representing its class. They are very interpretable as the flow of rules that can be used to classify the class can be represented as a tree. The individual DTs are appealing because they are easy to use and can be used with both numerical and categorical data. They can, however, get stuck in the process of overfitting, particularly with more complex datasets. Furthermore, DTs are the basic constituents of higher level ensemble methods like Random Forests (RFs). A comparison of the performance of a single DT with ensemble DT, helps to understand the advantages of bagging and random feature selection. DTs are also recently applied in various studies on ADHD prediction,

especially in applications that require an interpretable decision making process [36].

Mathematical Formulation of DT:

A Decision Tree recursively partitions the feature space by selecting, at each internal node, the feature and threshold that produce the purest possible child nodes. The default splitting criterion is the *Gini Impurity*, which for a node t containing samples from K classes is:

$$\text{Gini}(t) \;=\; 1 \;-\; \sum_{c=1}^{K} p_c^2, \tag{3.21}$$

where p_c denotes the proportion of samples of class c at node t. The Gini Impurity equals 0 when the node is pure (all samples belong to the same class) and increases as the classes become more mixed; for the binary case considered in this study ($K = 2$), the maximum value is 0.5.

When a node t is split on feature j with threshold s, the data $\mathcal{D}_t$ at that node is partitioned into a left child $\mathcal{D}_L = \{\, \mathbf{x} \in \mathcal{D}_t : x_j \le s \,\}$ and a right child $\mathcal{D}_R = \{\, \mathbf{x} \in \mathcal{D}_t : x_j > s \,\}$. The quality of the candidate split is measured by the weighted Gini Impurity of the resulting children, and the algorithm chooses the optimal pair (j^*, s^*) that minimizes

$$G(t, j, s) \;=\; \frac{|\mathcal{D}_L|}{|\mathcal{D}_t|}\,\text{Gini}(\mathcal{D}_L) \;+\; \frac{|\mathcal{D}_R|}{|\mathcal{D}_t|}\,\text{Gini}(\mathcal{D}_R). \tag{3.22}$$

The information gain achieved by the split is therefore

$$\Delta\text{Gini}(t, j, s) \;=\; \text{Gini}(t) \;-\; G(t, j, s). \tag{3.23}$$

Splitting continues recursively on each child node until a stopping criterion is satisfied, for example, until a leaf is pure (Gini = 0), the maximum tree depth is reached, or the number of samples at a node falls below a preset minimum. For an unseen sample $\mathbf{x}$, the predicted class corresponds to the majority class in the leaf node $\ell(\mathbf{x})$ to which $\mathbf{x}$ is assigned by the sequence of decision rules:

$$\hat{y} \;=\; \arg\max_{c \in \{0,\,1\}} \; p_c(\ell(\mathbf{x})), \tag{3.24}$$

where $p_c(\ell(\mathbf{x}))$ is the proportion of training samples of class c that fall into the leaf $\ell(\mathbf{x})$.

As an alternative splitting criterion, the *Entropy* measure can also be used:

$$H(t) \;=\; -\sum_{c=1}^{K} p_c \,\log_2(p_c), \tag{3.25}$$

with the corresponding information gain

$$\mathrm{IG}(t, j, s) \;=\; H(t) \;-\; \left[\frac{|\mathcal{D}_L|}{|\mathcal{D}_t|} H(\mathcal{D}_L) \;+\; \frac{|\mathcal{D}_R|}{|\mathcal{D}_t|} H(\mathcal{D}_R) \right].$$

The default Gini impurity criterion was kept as it is more computationally efficient than the entropy, and typically has similar performance in most binary classification problems. The DT model was chosen to compare with because of its high interpretability. The resulting trie can be seen as a flowchart of clinically relevant decision rules which can be especially useful in a healthcare context for ADHD risk screening.

These were trained on the Google Colabs cloud computing service. Something like complex algorithms and hyperparameter tuning which would otherwise take a lot of processing power on local workstations would have a better run in such an environment. The grid search control was utilized as Hyperparameter tuning with the aim of maximizing the performance of a model. In this process five-fold cross-validation was employed in the training set to ensure high model validation and improved generalization. This method was used to divide the training data into five equal folds. Each iteration was done with four folds during training and one-fold during validation. This process was repeated five times in order to use each fold as the validation set at least once. To reduce variation and ensure that the overall performance is not biased towards a specific train-validation-split, the final performance was obtained through an average of all the folds. Hyperparameters are foreign configuration parameters: the number of trees in a RF, the regularization parameter in SVM or LR or the maximum depth of a DT, are parameters that must be specified in advance but not learned during training. The grid search CV systematically searches a predefined set of combinations of hyperparameters. The model is trained and evaluated by five-fold cross validation on each of the combinations. This method includes the training dataset (88 people, or 80 per cent of the total data) segregated into five-fold folds. The model is trained in four folds and the lastfold is used to validate a model. To ensure that each fold provides the validation set with the correct frequency, this procedure is repeated five times. The average performance over such five cycles provides more reliable estimate on the performance of the model and helps in selecting the best hyperparameter configuration that gives the best generalization capabilities. This carefully selected optimization process optimizes the predictive ability of the models to assess the risk of ADHD and ensures that not only are the models well-trained but they are also configured to perform optimally on unseen data.

3.6 Model Evaluation

After the training and hyperparameter optimization phase was completed, every machine learning model was assessed to establish its capabilities in predicting unknown data and generalization. In a bid to provide a dispassionate assessment of the suitability of the models to work under real-world conditions, this important step relied on the special 20 percent test set which was not part of the training process. The study was based on a number of popular classification metrics. A Confusion Matrix is a performance evaluation tool that is employed in the classification problem to study the accuracy of a machine learning model in predicting various classes. It compares the actual values and the predicted ones generated by the model. There are four important parameters in the confusion matrix: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These parameters are further used to calculate various evaluation metrics to measure the overall performance of the proposed model, which includes Accuracy, Specificity, Precision, Recall and F1-Score. The performance metrics were defined as follows:

- Accuracy: shows the percentage of correctly classified cases (true positives and true negatives) relative to the total number of cases. It shows how accurate the model is overall.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.26}$$

- Specificity: calculates the percentage of real negative cases that are accurately classified as negative. It represents the model's capacity to accurately identify people without ADHD in the context of ADHD prediction

$$Specificity = \frac{TN}{TN + FP} \tag{3.27}$$

- Recall: calculates the percentage of real positive cases that are accurately classified as positive. In terms of ADHD prediction, it shows how well the model can identify those who actually have ADHD.

$$Recall = \frac{TP}{TP + FN} \tag{3.28}$$

- Precision: calculates the percentage of positive identifications that were true. It shows how well the model avoids making misleading positive predictions.

$$Precision = \frac{TP}{TP + FP} \tag{3.29}$$

- F1-Score: A single measure that strikes a balance between precision and recall is the harmonic mean of the two. Since it penalizes models that prefer one class over another, it is very helpful when working with imbalanced datasets.

$$F1 - Score = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{3.30}$$

Area Under the Receiver Operating Characteristic Curve: The AUC is used to evaluate the total discriminating ability of a classification model given all possible levels of classification. The larger the ability to separate positive and negative classes is indicated by a higher AUC value (nearer to 1.0), which indicates a better trade off between sensitivity and specificity with different decision limits. An AUC of 1.0 indicates perfect discrimination and an AUC of 0.5 indicates that there is no discrimination at all (it is the same as random guessing).

Chapter 4

App Interface

The final step in this project is the development of the **SHUCHOK** application interface, which is not limited to developing model. This interface takes complex forecasts and converts them into an easy-to-use tool, which is the practical endpoint of the machine learning model. The application, designed to meet the special needs of the Bangladeshi context, tries to eliminate barriers to early ADHD diagnosis by providing a bilingual, culturally sensitive, and easy-to-use platform on which to perform the initial risk assessment.

4.1 SHUCHOK Overview

The primary objective of the proposed application interface is to be used as a bilingual self-assessment measure of the risk of ADHD in Bangladeshi children. This parent-educator-guardian friendly interface tries to bridge the vast knowledge and access gap in terms of affordable ADHD screening services in the region. The app provides an estimated level of ADHD risk by allowing caregivers to input the observed behavioral responses into a questionnaire format; it is an initial indicator and not a final diagnosis. The primary two screens that comprise the application workflow are survey input screens and risk output screens which are designed to be easy to use and intuitive. The users will give their observations on the behavior of a child by setting through a series of questions which are translated into Bengali and adapted based on the DSM 5 [5] criteria. An output risk page is displayed with the result of the processed data after the underlying machine learning model has processed the received data. The design of this interface gives much attention to usability. The interface integrates various key design principles to ensure its broad accessibility and ease of adoption particularly in a diverse user base. It is characterized

by a simple, easy to use design that reduces the amount of complexity and reduces the cognitive burden on users. It also uses kid-friendly colors and icons in its graphic design to ensure a friendly and engaging experience that will soften the stigma that often accompanies neurodevelopmental issues. Bilingual support is necessary to be able to provide the tool to a bigger community, despite their degree of English proficiency, to reflect the linguistic diversity of Bangladesh. To enhance participation in preliminary ADHD risk assessment, the following strategy design aims at making the process of self-assessment friendly and non-threatening.

4.2 Interface Design

Consistent with its purpose as an awareness and self-assessment tool for parents and teachers, the interface — prototyped in Canva — prioritises clarity, usability, and visual appeal. To establish a non-clinical, welcoming environment, the overall visual language will be based on a combination of kid-friendly colors (light blues, greens, yellows, and oranges) and simple layout and minimalist elements. This design choice is aimed at reducing the stigma and anxiety often linked with mental health assessment and enhances the comfort of users and encourages honest involvement. The navigation system is simplified to the maximum, and there are no complex lists with lots of complex buttons such as start assessment and next. Every screen of the interface is a step-by-step screen-based flow that focuses on one activity at a time, and it is easy and understandable to users to complete the process. The main screens layout is shown as follows:

Welcome/Introduction Screen: The initial screen describes the purpose of the application that is to provide an initial risk assessment of ADHD. It has a button to start assessment and offers language options (English/Bengali) to ensure that it can be used by a high number of users. The UI conveys the idea that the tool is not a diagnostic system and creates a positive atmosphere.

User Identification Screen: When the assessment has started, the user is asked to select his role (e.g., parent or teacher). This stage enhances reliability of the data collected and ensures that answers are contextual.

Student Information Input Screen: Through this screen, the child age and sex among other simplistic demography are collected. Simple input fields and clear labelling make it easy to enter data even to users with different levels of technical knowledge.

Survey Input Screens: These screens are the key to the method of data collection. To maintain focus and reduce the mental burden, only one behavioral question is shown at a

time on each panel. They are asked with large, readable fonts and are mostly in Bengali with an optional English support to make them more accessible. They are founded on the DSM-5 ADHD criteria.

- Question Display: Questions are in large and legible text, and are available. This is a dual language display that will guarantee real time understanding among a wide range of users.
- Input Fields of Behavioral Responses: "Never", "Sometimes", "Often" and "Always" are the examples of the response options which are organized and presented in a logical and comprehensible manner. These options are easy to select, particularly to the consumers who have no previous experience in the digital world, since they are displayed as large and clickable selections.
- Progress Indicator: A progress bar or numbering system on the top of the screen allows users to know where they are in the questionnaire, to manage expectations and to reduce fatigue.
- Navigation Buttons: To guide the user on the evaluation, there are large Next buttons on the screens. The design ensures smooth and easy user experience to the point of submission by eliminating unnecessary backwards complication.

Risk Output Screen: This screen is intended to present the predicted ADHD risk assessment clearly and instantly.

- Visual Risk Indicator: This is represented by a clear visual representation of danger, using a Low–Moderate-High meter and intuitive color coding (green Low, yellow Moderate, red High). This removes the technical skills and the result can be understood instantly by the users.
- Descriptive Text: A short description of the expected amount of risk is provided under the graphic signal. The wording underlines that the result must not be considered a medical diagnosis and makes it quite obvious that it is an initial examination.
- Guidance and Resources: The result screen will prompt users to visit with trained medical experts such as pediatricians or child psychologists and provide practical next steps, particularly with moderate and high-risk cases.

The unified visual code, combined with the effective interaction features and a bilingual support, will help to make the interface not only enjoyable to look at but also quite practical and usable, thus encouraging its wide adoption as a first-line screening tool in Bangladeshi setting. The prototype design of the proposed mobile application interface, designed to facilitate easy ADHD screening, is depicted in Figure 4.1

4.3 Prototype

The application interface prototype demonstrates the survey as an input to the risk assessment as an output to give a tangible representation of what the user will experience. A major element of this prototype is the Low-Moderate-High Meter that is meant to provide an immediate and understandable indication of the risk level to expect in ADHD. The meter splits the output of the machine learning model into three levels of risk graphically:

Low Risk: This is the level that is typically denoted by a green segment on the meter and one that shows the fact that the behavioral and demographic profile of the child as identified through the questionnaire is more likely to be similar to the trends observed in individuals without ADHD. This is very comforting but carries with it a caution that the instrument is not diagnostic and that further observation is recommended.

Moderate Risk: A yellow/amber segment indicates that the child's profile places them in an intermediate category — some ADHD-like behaviours may be present but are either not numerous enough, or are inconsistent, to suggest high risk. The tool recommends closer observation and discussion with a paediatrician or child specialist.

High Risk: This level is identified by a red segment meaning that the profile of the child is very similar to the patterns associated with ADHD. The presence of some interesting behavioral and demographic features indicates a higher risk of ADHD. This is one of the results that are clearly highlighted in the prototype and an immediate recommendation on immediate professional consultation and a proper diagnostic analysis by a licensed health specialist.

The careful design of the Display of Results panel makes it possible to present this important information in a responsible and clear way. On this screen, the expected danger level is graphically presented by the meter when the user has entered his/her survey responses. In order not to overload consumers with complex statistical information and to give a more intuitive interpretation, it is essential that the numerical probability generated by the machine learning model should be transformed into the following category risk levels. Alongside the visual meter, the display of the results contains:

- Predicted ADHD Risk: Moderate - An example of a succinct sentence that expressly presents the result is moderate.
- Observe: An obvious warning to indicate that the app is not a diagnostic tool and only provides an initial evaluation. It stresses that the formal diagnosis can be made only by a licensed medical practitioner.
- Practical Next Steps: The prototype includes elaborate guidelines depending on the

level of risk. There should be a professional assessment by a pediatrician, a child psychologist or a child psychiatrist in case the risk is moderate or high. A full-fledged version could also contain contact information or links to related local resources (like ADHD support groups and mental health organizations in Bangladesh).

The project goal of enabling early detection and intervention is reached by the prototype focus on the simplicity of the meter with color codes and a set of actionable and easy to follow recommendations that ensure the user, irrespective of the background, can understand the consequences of the test and is guided to the necessary follow-up measures.

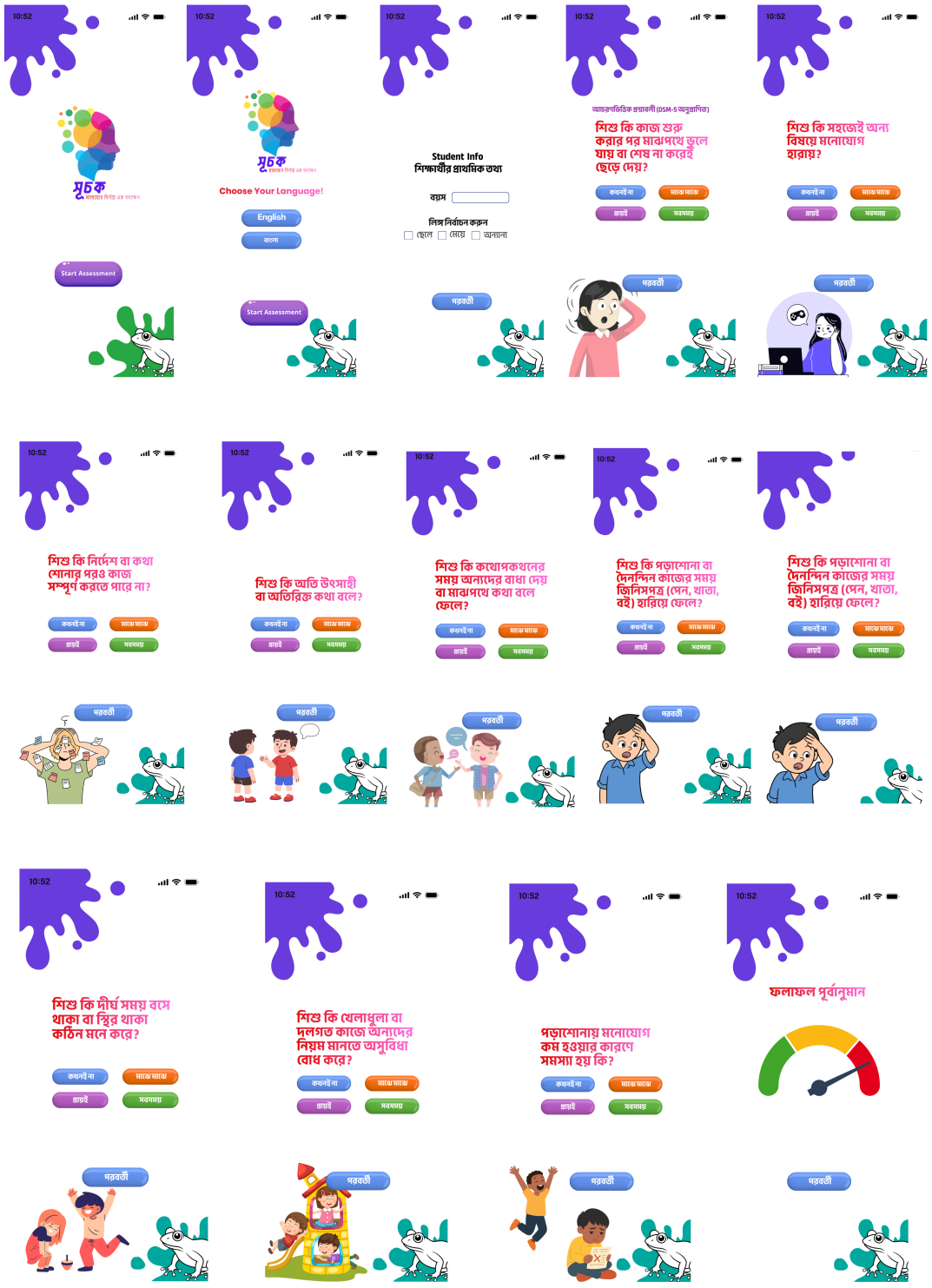

Figure 4.1: SHUCHOK-ADHD Risk Assessment Mobile Application (User Interface Design)

Chapter 5

Results and Analysis

This section discusses the effectiveness of the machine learning models that have been developed to predict ADHD and their interpretability and importance of identified risk factors. The assessment utilizes valuable performance measures, visual analytics such as confusion matrices and ROC curves, and a careful analysis of the importance of features to confirm the effectiveness of the proposed methodology.

5.1 Performance Analysis of Random Forest Model

Random Forest (RF) classifier was chosen as one of the major ensemble learning models because it is less prone to overfitting and can deal with non-linear relations. A 5-fold cross validation protocol was set up to evaluate it statistically. This method divides the data set into 5 different subsets, each of which has every data point used as a training and testing example over the different experimental cycles.

5.1.1 Error Analysis via Confusion Matrices

The internal classification logic of the RF model is visualized through the confusion matrices in Figure 5.1, labeled (a) through (e). These matrices allow for a direct observation of Type I (False Positive) and Type II (False Negative) errors.

- **Class Stability**: The model has been able to classify most of the cases consistently across all folds. It is easy to see that the Fold 2 (b) and Fold 3 (c) are close to ideal situations. Fold 2 had a 0% False Positive and Fold 3 had a 0% False Negative,

meaning that the model had 100% sensitivity for the data partition in question.

- **Error Distribution**: The misclassifications are very small and usually only one per classification. In particular, Folds 1, 3, 4 and 5 only had one False Positive (when it is Class 0, the model predicted Class 1). This indicates that the Random Forest model is a bit more conservative in predicting the positive class, focusing on accuracy to prevent false alarms.

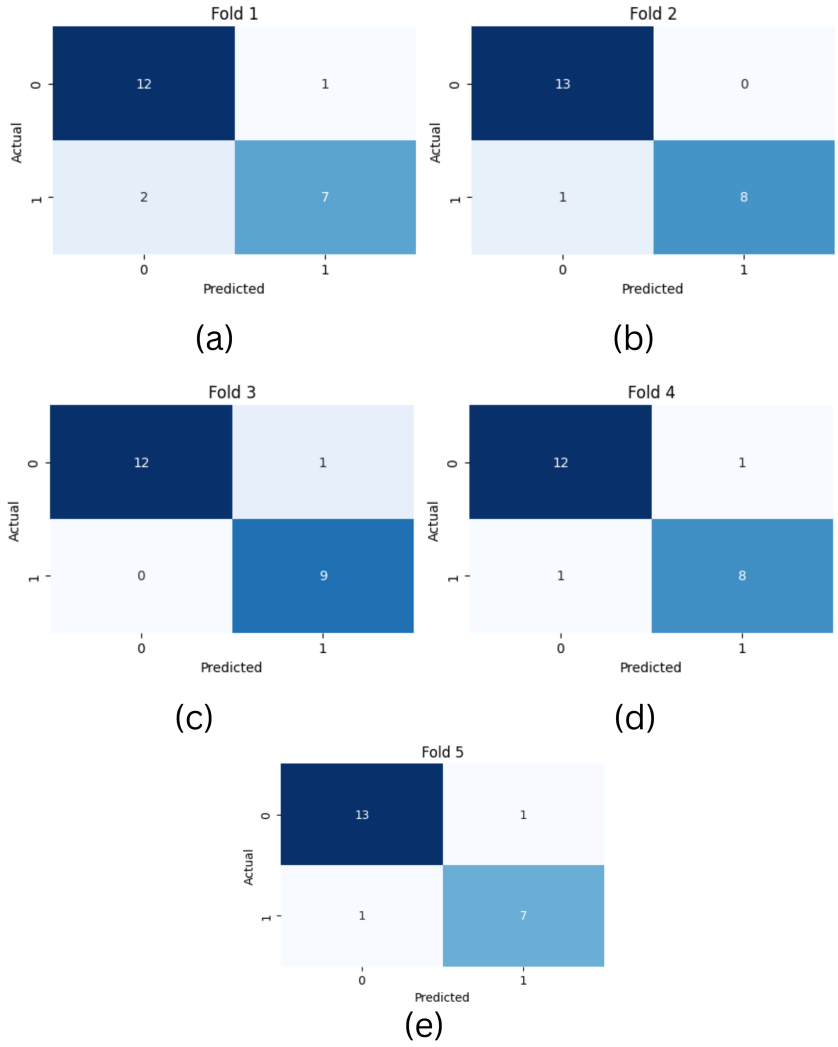

Figure 5.1: 5 Fold Confusion Matrices for RF

5.1.2 Statistical Evaluation of Performance Metrices

A comparison of quantitative effectiveness of the RF model is presented in Table 5.1. This table shows an overview of model metrics from all validation folds.

The average value of Accuracy obtained by the model was 91.82%, which indicated the high overall power of prediction. For certain contexts, it is important to note that ac-

Table 5.1: Random Forest - Detailed Evaluation Metrics per Fold

Fold	Accuracy	Recall	Specificity	Precision	F1-Score	AUC
1	0.8636	0.7778	0.9231	0.8750	0.8235	0.9402
2	0.9545	0.8889	1.0000	1.0000	0.9412	0.9957
3	0.9545	1.0000	0.9231	0.9000	0.9474	0.9701
4	0.9091	0.8889	0.9231	0.8889	0.8889	0.9615
5	0.9091	0.8750	0.9286	0.8750	0.8750	0.9062
Average	**0.9182**	**0.8861**	**0.9396**	**0.9078**	**0.8952**	**0.9548**

curacy can be misleading, and the Average F1-Score (89.52%) is a more comprehensive measure of the balanced performance of the model in terms of Precision and Recall. The Average Specificity is 93.96% indicating that the model is very good at excluding the negative cases (Class 0), thus avoiding too many interventions or false detections.

5.1.3 Probabilistic Interpretation (ROC/AUC)

To evaluate the model's discriminative threshold, the Receiver Operating Characteristic (ROC) analysis was conducted, as shown in Figure 5.2.

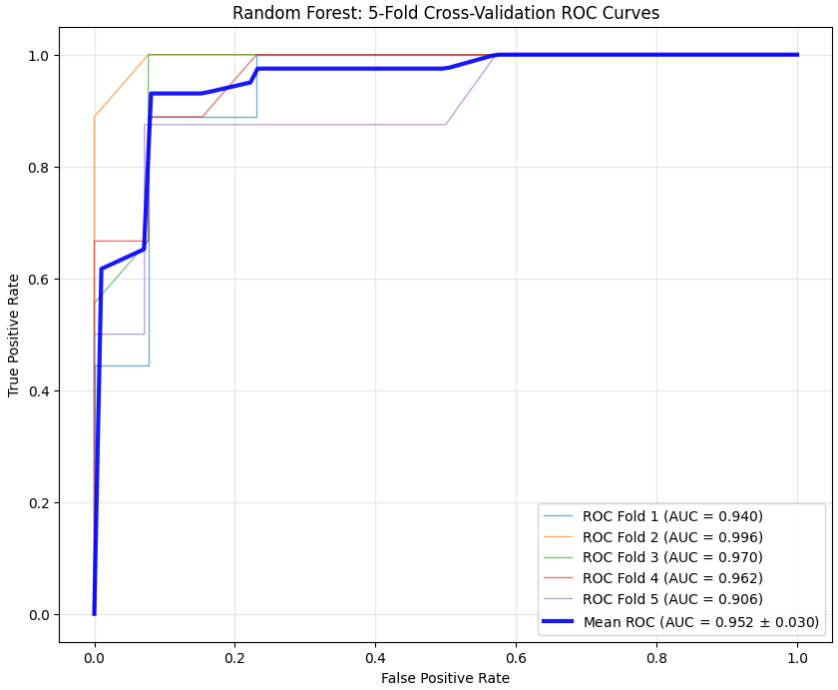

Figure 5.2: Random Forest ROC Curve

- **Discriminatory Power**: The model could be presented with a Mean AUC of 0.952, having a standard deviation of ± 0.030. With an extremely high AUC value, there is a 95.2% chance that the model will be able to distinguish between a randomly

selected positive instance and a randomly selected negative instance.

- **Consistency**: All the curves shown in Figure 5.2 appear to be pointing towards the top left corner, indicating high consistency. The small variance (0.030) implies that the model is not very sensitive to the noise in each fold, and is thus a promising model for transfer to out-of-sample data.

5.2 Performance Analysis of Support Vector Machine (SVM)

To explore its ability in finding an optimal hyperplane to separate the classes, the Support Vector Machine (SVM) classifier was implemented. Rigorously tested performance of the model was done via 5-fold cross validation to show that its predictive success varies considerably with respect to the way the data was split.

5.2.1 Detailed Error Analysis (Confusion Matrices)

The classification performance for each fold is visualized through the confusion matrices in Figure 5.3 labeled (a) through (e).

- **Fold Variation**: As opposed to the ensemble method, the SVM model was more sensitive to the data distribution in particular folds. Fold 2 (b) and Fold 5 (e) proved to be very successful with Fold 2 having almost perfect classification (only one False Negative) and Fold 5 with only four False Negatives.
- **Misclassification Trends**: The highest challenge for the SVM was the Fold 1 (a): 3 False Positive and 4 False Negative. This suggests that in some subsets, SVM kernel classification was not as clear-cut as other classifiers, resulting in a more even distribution of Type I and Type II errors.
- **Class 0 Proficiency**: The SVM in general had higher accuracy for the negative class (Class 0) as occurred in the RF model in Folds 2, 3 and 5.

5.3 Quantitative Statistical Evaluation

The performance metrics across the 5-fold cross-validation are summarized in Table 5.2.

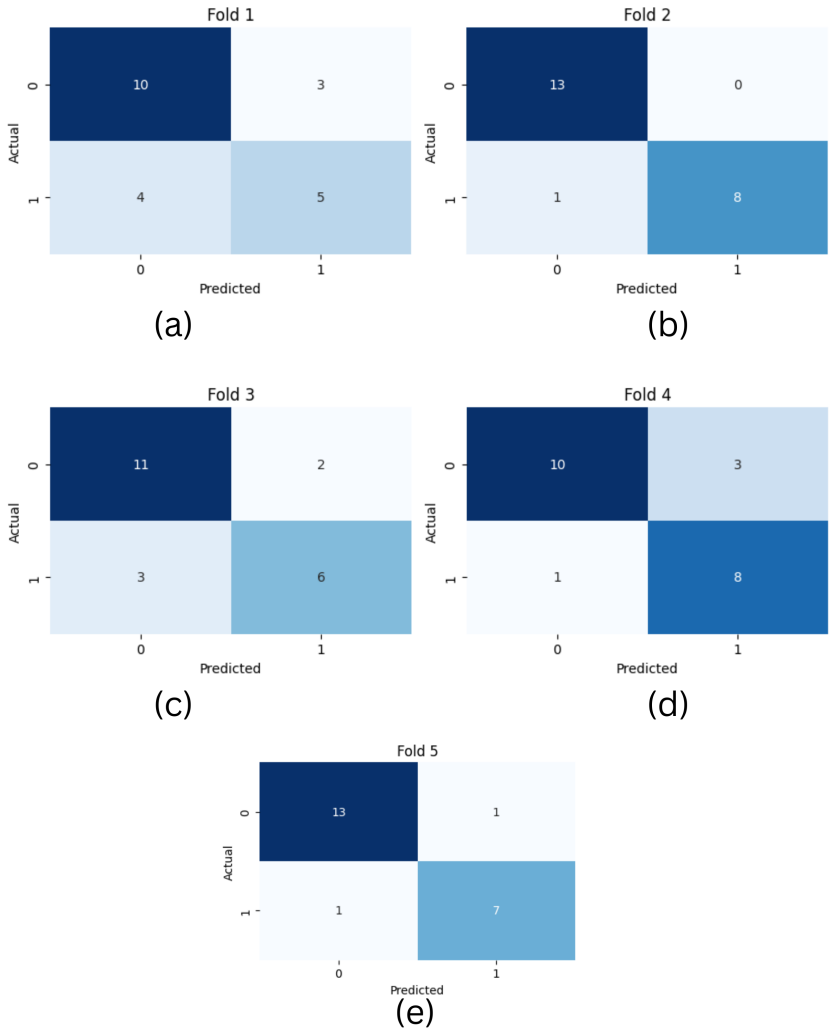

Figure 5.3: 5 Fold Confusion Matrices for SVM

Table 5.2: SVM - Detailed Evaluation Metrices per Fold

Fold	Accuracy	Recall	Specificity	Precision	F1-Score	AUC
1	0.6818	0.5556	0.7692	0.6250	0.5882	0.7692
2	0.9545	0.8889	1.0000	1.0000	0.9412	0.9915
3	0.7727	0.6667	0.8462	0.7500	0.7059	0.8974
4	0.8182	0.8889	0.7692	0.7273	0.8000	0.8376
5	0.9091	0.8750	0.9286	0.8750	0.8750	0.9375
Average	**0.8273**	**0.7750**	**0.8626**	**0.7955**	**0.7821**	**0.8866**

The average accuracy of SVM model is 82.73%. Average Recall (77.50%) and Precision (79.55%) were respectable values though, but much lower than the Random Forest model and yielded an average F1-Score of 78.21%. The difference between the best performing fold and the worst performing fold (95.45% vs 68.18%) indicates that SVM can be very effective, but could possibly need a more carefully tuned hyperparameter or feature scaling in order to achieve consistent stability across all portions of the data.

5.3.1 ROC Curve and Discriminatory Analysis

The diagnostic ability of the SVM is further analyzed via the ROC curves presented in Figure 5.4

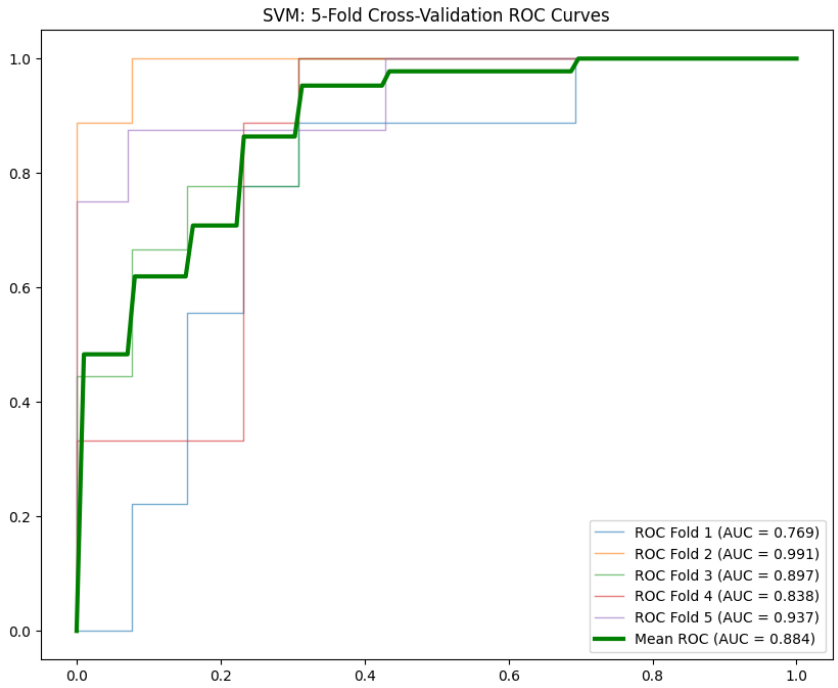

Figure 5.4: ROC Curve for SVM

- **AUC Metrics**: The Mean AUC of this model was 0.8866. This corresponds to a

"good" to "excellent" level of discrimination, and the visual spread of the individual fold curves in Figure 5.4 verifies the fluctuation in performance mentioned in the "Metrics" table.

- **Threshold Performance**: Fold 2 had an outstanding AUC of 0.991, almost reaching the top left corner of the graph. It is worth noting, however, that the lower AUC of Fold 1 (0.769) gives the impression that the model is dependent on the characteristics of the training fold.

5.4 Performance Analysis of Logistic Regression

The LR model was used as a baseline classifier for the purpose of determining whether the data is linearly separable. It was tested by the 5-fold cross validation method as a benchmark to compare with more complex non-linear models such as SVM and ensemble methods such as Random Forest.

5.4.1 Error Analysis and Classification Distribution

The classification behavior of the LR model across all folds is illustrated in the confusion matrices in Figure 5.5, labeled (a) through (e).

- **Consistency in Classification**: The model's error distribution was stable, with 1-3 errors per fold for each fold. The best performance for this model was obtained with Fold 3 (c) which resulted in one False Positive and one False Negative.
- **Error Bias**: Model False Negatives (actual 1s predicted as 0s) were greater for Folds 1, 2, and 4 than Folds 3 or 5. This means that Logistic Regression's linear decision boundary can sometimes fail to pick up positive examples where the feature space is somewhat similar.
- **True Negative Performance**: The model continued to be very successful at classification, correctly classifying 10 - 12 negative examples in each fold of the test.

5.4.2 Statistical Evaluation of Quantitative Metrics

The detailed performance indicators for the LR model are presented in Table 5.3.

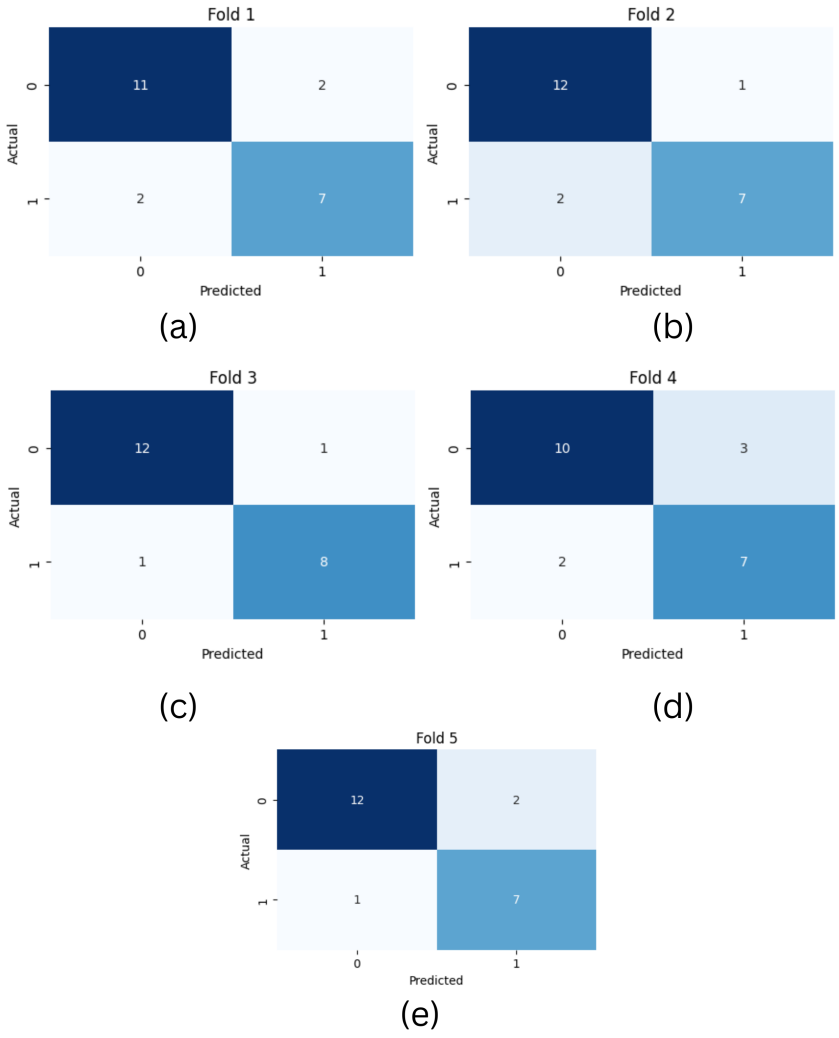

Figure 5.5: 5 Fold Confusion Metrices of LR

Table 5.3: Logistic Regression - Detailed Evaluation Metrices per Fold

Fold	Accuracy	Recall	Specificity	Precision	F1-Score	AUC
1	0.8182	0.7778	0.8462	0.7778	0.7778	0.8376
2	0.8636	0.7778	0.9231	0.8750	0.8235	0.9658
3	0.9091	0.8889	0.9231	0.8889	0.8889	0.9658
4	0.7727	0.7778	0.7692	0.7000	0.7368	0.9060
5	0.8636	0.8750	0.8571	0.7778	0.8235	0.9018
Average	**0.8455**	**0.8194**	**0.8637**	**0.8039**	**0.8101**	**0.9154**

Discussion of Mean Results: The LR model has a decent average Accuracy of 84.55%. It achieved an accuracy greater than that of the SVM model (82.73%), but it was less than the Random Forest model (91.82%). The average F1-Score of 81.01% represents a good precision and sensitivity. Interestingly, the model had a lower variation between the best and worst accuracy folds than the SVM, indicating that the linear approach is pretty reliable for this data set.

5.4.3 AUC and ROC Curve Interpretation

The discriminatory power of the LR model is visualized through the ROC curves in Figure 5.6

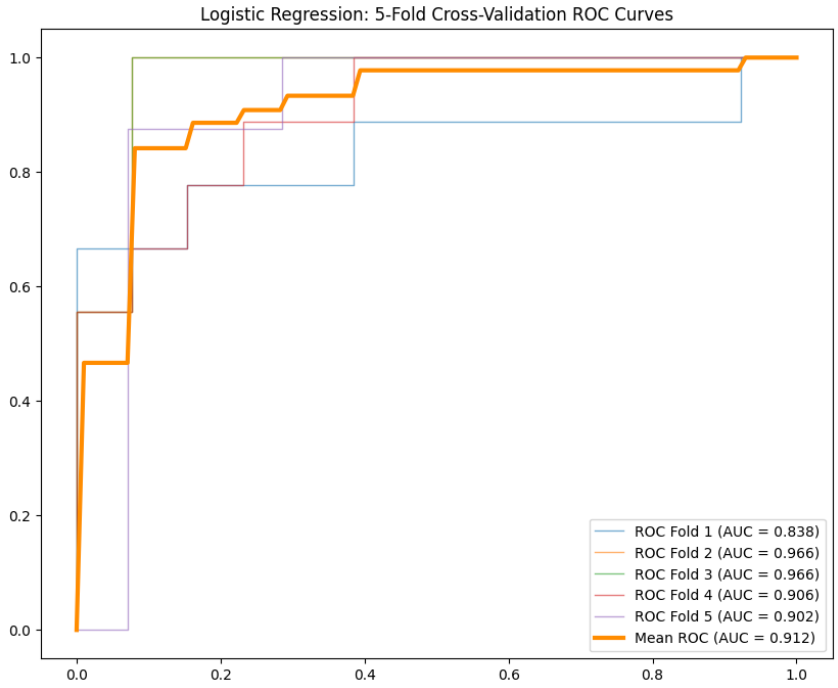

Figure 5.6: ROC Curve for LR

- **Overall Performance**: The Mean AUC of the model was 0.912. This value represents a strong confidence level that the model can correctly classify the two classes for a range of decision thresholds.
- **High-Performing Folds**: Folds 2 and 3 both achieved an impressive AUC of 0.966, as shown by the curves reaching significantly toward the upper-left corner of the plot.
- **Clinical/Practical Reliability**: The model also had an AUC of 0.838 even at its worst performance (Fold 1). This overall shape of the Mean ROC (shown as the orange line) is an indication that Logistic Regression is a good, low-complexity candidate for this classification problem.

5.5 Performance Analysis of Decision Tree

The Decision Tree (DT) classifier was tested to witness the performance of a non-linear classification model with the data set. The Decision Trees are extremely interpretable, but in terms of the 5-fold cross validation, the model showed the largest variance of performance among all of the classifiers tested.

5.5.1 Analysis of Classification Errors (Confusion Matrices)

The classification performance across the experimental folds is visualized in the confusion matrices in Figure 5.7, labeled (a) through (e).

- **Sensitivity to Data Splitting**: There was a lot of variation in the predictive accuracy of this model. The Recall for Fold 1 (a) was 1.0 with all positive instances correctly identified while the Recall for Fold 5 (e) was not very good as it misclassified 4 out of 8 positive instances (FN=4).
- **False Positive Trends**: Model was always making False Positives in all folds, with maximum False Positive (FP=4) in Fold 4. This indicates that model decision nodes are possibly overfitting to some noise or outliers in the training subsets.
- **Predictive Instability**: The DT has a high variance with small-to-medium size datasets, such as when going from 100% Recall in Fold 1 to 50% Recall in Fold 5.

5.5.2 Quantitative Metrics and Statistical Summary

The performance metrics for the Decision Tree are detailed in Table 5.4.

Table 5.4: Decision Tree - Detailed Evaluation Metrics per Fold

Fold	Accuracy	Recall	Specificity	Precision	F1-Score	AUC
1	0.8636	1.0000	0.7692	0.7500	0.8571	0.8846
2	0.8182	0.7778	0.8462	0.7778	0.7778	0.8120
3	0.8182	0.6667	0.9231	0.8571	0.7500	0.7949
4	0.7273	0.7778	0.6923	0.6364	0.7000	0.7350
5	0.7273	0.5000	0.8571	0.6667	0.5714	0.6786
Average	**0.7909**	**0.7444**	**0.8176**	**0.7376**	**0.7313**	**0.7810**

Discussion of Mean Results The lowest accuracy among the four models evaluated was that of the Decision Tree with an average of 79.09%. The average F1-Score = 73.13%

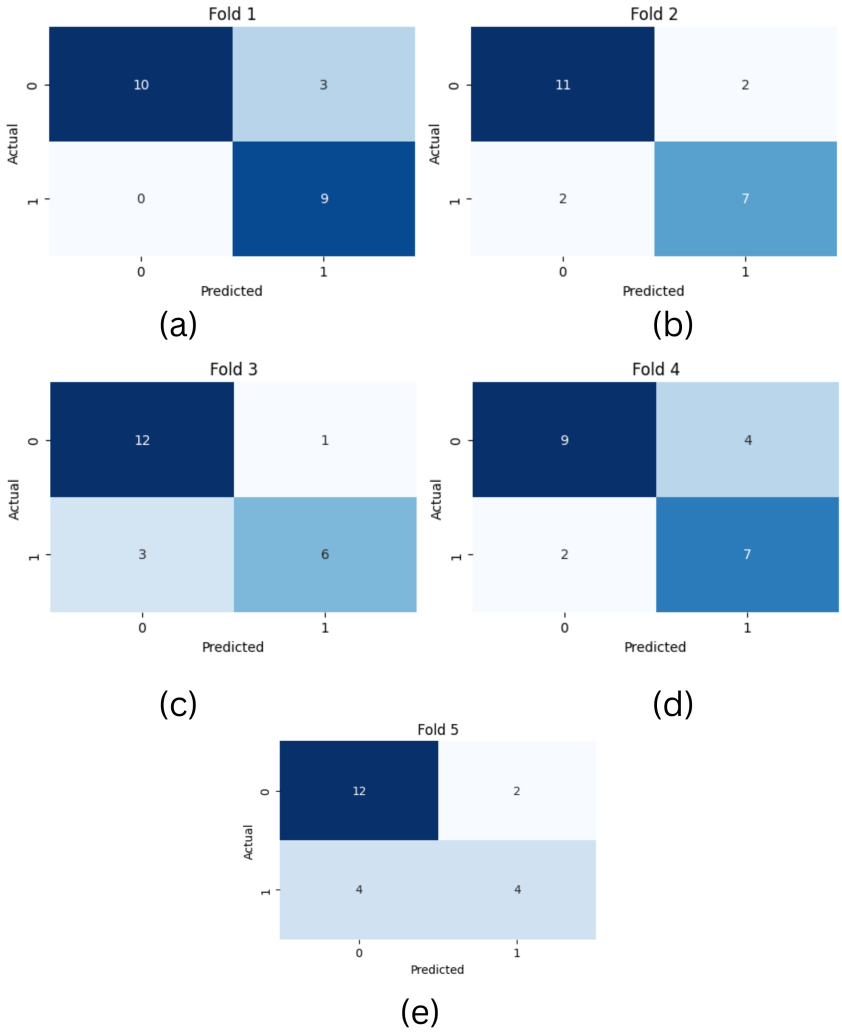

Figure 5.7: 5 Fold Confusion Metrics of DT

also indicates that the model tends to have weaker performance in achieving a balance between Precision and Recall. Comparing the accuracy of the single tree model with the accuracy of the ensemble approach (Random Forest) in Folds 4 and 5 (72.73% accuracy) shows that the single tree model is not as robust.

5.5.3 ROC Curve and Discriminatory Analysis

The diagnostic thresholding of the Decision Tree is visualized in Figure 5.8

- **AUC Performance**: The average AUC of this model was 0.781. The image provided in Figure 5.8 has "stair-step" patterns, rather than smooth curves, which is typical of the decision boundaries formed by a single tree.
- **Fold-Level AUC**: The discriminatory power decreases from Fold 1 (AUC=0.885) to Fold 5 (AUC=0.679). The AUC of 0.679 is considerably below the performance of the other models, which indicates that the Decision Tree is not the most robust model for this particular classification problem when compared to the other models, unless some additional pruning or boosting methods are used.

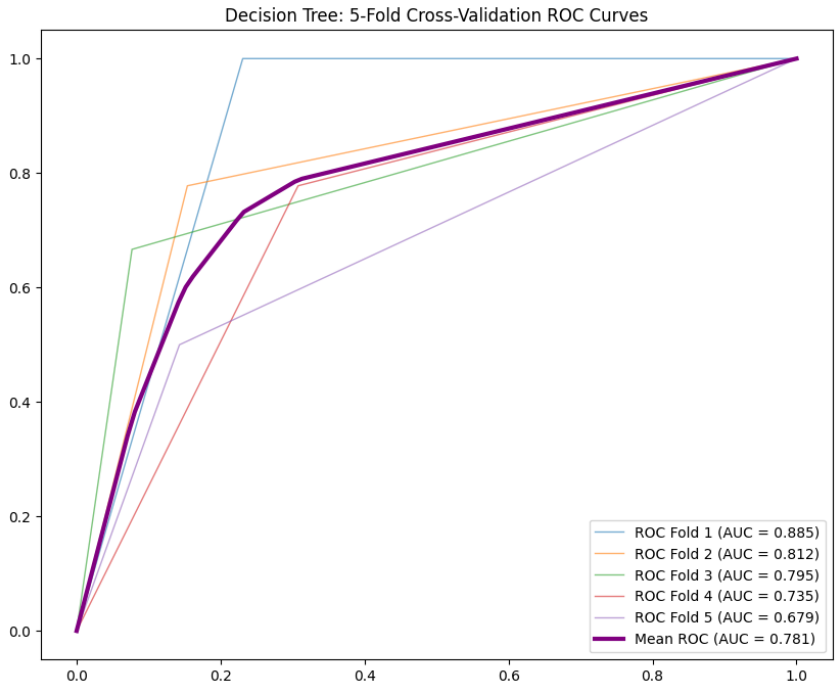

Figure 5.8: ROC Curve for DT

5.6 Comparative Analysis of All Models

After each classifier was evaluated independently, a comparison was made in order to decide which model would be the best in this classification problem. The average result obtained from the five folds can be used to determine which algorithms have the most predictive power, are most stable, and have the best class-discrimination ability.

5.6.1 Final Cross-Validation Summary

The holistic comparison of the four models—Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), and Decision Tree (DT)—is summarized in Table 5.5.

Table 5.5: Comparison Summary Of All Model Performance

FINAL CROSS-VALIDATION SUMMARY (All Models Compared)						
Model	**Avg AUC**	**Avg Recall**	**Avg Spec.**	**Avg Acc.**	**Avg Prec.**	**Avg F1**
Random Forest	**0.9548**	**0.8861**	**0.9396**	**0.9182**	**0.9078**	**0.8952**
Logistic Regression	0.9154	0.8194	0.8637	0.8455	0.8039	0.8101
SVM	0.8866	0.7750	0.8626	0.8273	0.7955	0.7821
Decision Tree	0.7810	0.7444	0.8176	0.7909	0.7376	0.7313

Key Findings from the Comparative Table

- **Superiority of Ensemble Learning**: The Random Forest model was the best in each of the evaluation metrics. It achieved an Average Accuracy of 91.82% and Average F1-Score of 89.52% which was significantly higher than the other models. It has a high precision (90.78%), meaning that it is very reliable when it makes a positive prediction.
- **The Baseline Robustness of LR**: Surprisingly, LR has got the second best position compared to the SVM and DT. With an Avg. It achieves an AUC score of 0.9154, indicating that features in this dataset are well suited to being represented by linear boundaries, and that linear boundaries provide a high accuracy but low computational cost way of modeling such relationships as compared to ensemble methods.
- **The Variance of SVM and DT**: Moderate results were obtained by the SVM, with an accuracy of 82.73% and a recall of 77.50%. The Decision Tree performed

the poorest of all the models on the Avg. category, and was also poor on all other categories. It is prone to noise and overfitting when used individually (when not in an ensemble) as indicated by its F1-Score (73.13%).

5.6.2 Multi-Model ROC Curve Analysis

To further visualize the performance hierarchy, the Comparison of Mean ROC Curves, shown in Figure 5.9, provides a clear representation of the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR) for all four models simultaneously.

- **Random Forest (Blue Line)**: The RF curve is also located closest to the top left corner of the plot than any of the other model curves. This visual ranking supports its statistical superiority, since it has the best sensitivity at the lowest false alarm rate for all thresholds it is possible to choose.
- **Logistic Regression (Orange Line)**: The LR curve is very close to the RF model. It steadily rises with the threshold value towards the top-left corner and therefore it is the second best classifier, with a significant and stable margin over SVM and DT in most part of the threshold range.
- **Support Vector Machine (Green Line)**: When plotted on an SVM curve, the curve has a "stepped" progression in that it is dependent on the support vectors. It does well at higher thresholds, but obviously lies below the RF and LR lines, meaning that it's more likely to be misclassified at more sensitive thresholds.
- **Decision Tree (Purple Line)**: The DT curve is the least performing in this comparison. It has an almost linear trajectory with very small discriminatory power, indicated by its angle of approximately 45 degrees. The DT curve clearly has a much larger portion of space between itself and the upper left-hand corner, and a lower Mean AUC of 0.7810 compared to the other models.

5.7 Final Model Selection and Justification

The RF model was chosen as the best model for the classification of ADHD in this study, following the detailed assessment. This choice is substantiated by several factors. First, the model consistently delivered the best performance for all performance metrics: accuracy, precision, recall, F1-score and AUC. Second, it showed excellent generalisation performance, with little variation between cross-validation folds. Third, the confusion

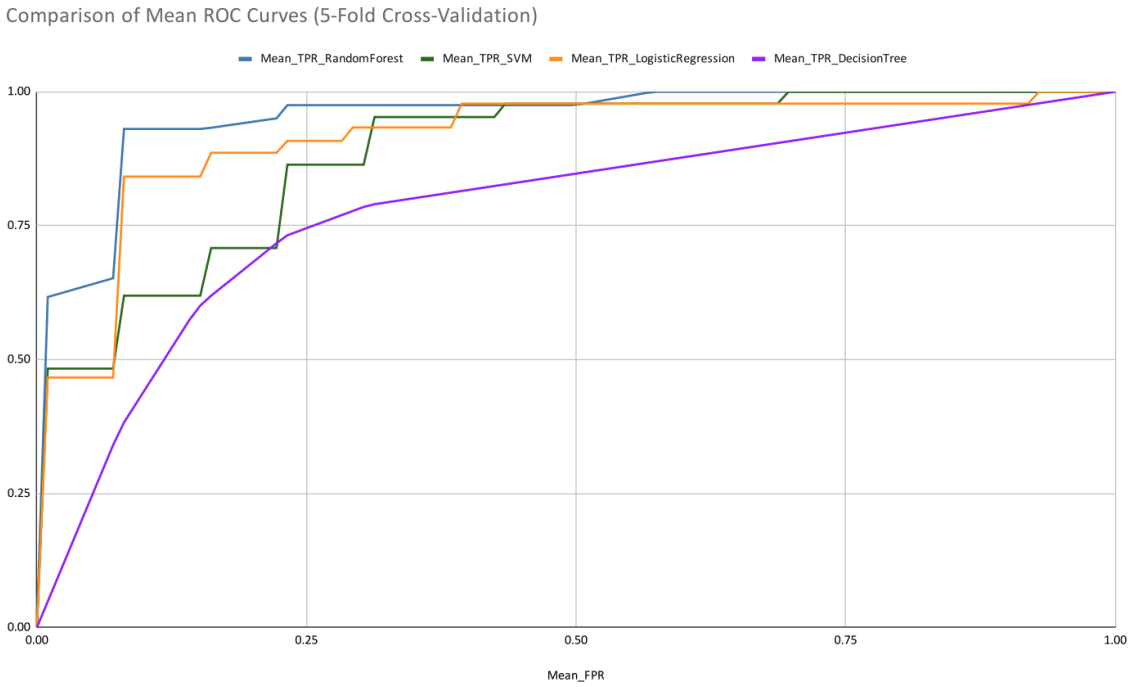

Figure 5.9: ROC Curve for Best Algorithm

matrix analysis showed its ability to accurately predict both ADHD and non-ADHD participants. Results of the classification performance of this model showed balanced prediction results in the confusion matrix. The overall confusion matrix from the 5-fold cross-validation reveals that the model successfully identified 54 negative samples and 33 positive samples in Figure 5.10, with lower misclassification rates for both false positives (12) and false negatives (11). This showed that the model was able to make an accurate distinction between ADHD related and non-ADHD risk cases with consistent predictions. The well-balanced class distribution and the low classification error exhibited ensures strength and generalization power of the RF model, making it the best one for the proposed system.

Moreover, the ROC analysis guaranteed that the AUC values are unbiased estimates of the true discriminative performance. This ensures the validity of the results and confirms RF as the optimum model. Overall, the **Random Forest** model provides the optimal combination of accuracy and reliability making it the preferred model for this classification problem.

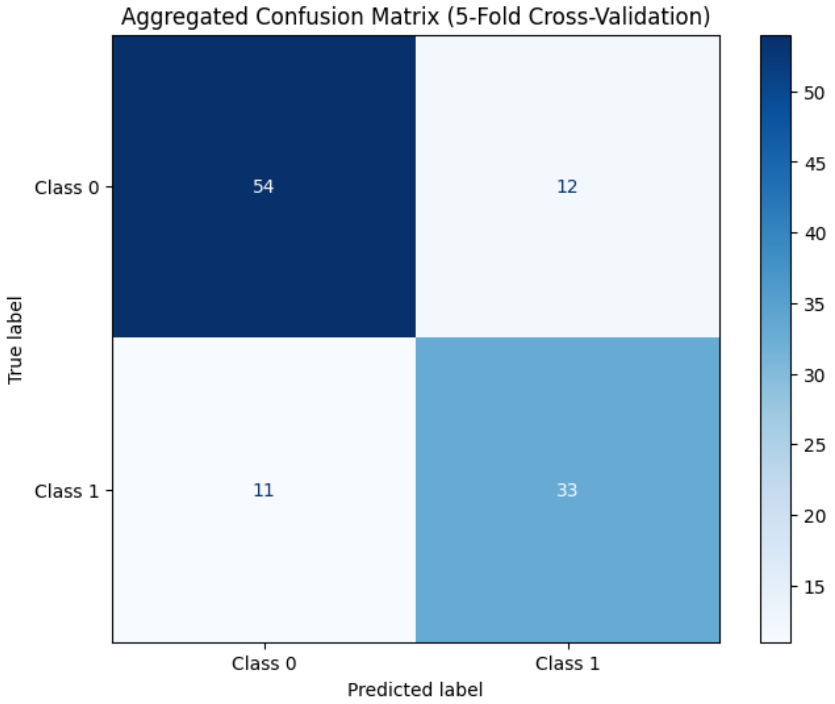

Figure 5.10: Combined Confusion Matrix for Best Algorithm

5.8 Feature Importance

To grasp a machine learning model's judgment and obtain a deeper understanding of the underlying phenomenon, it is essential to comprehend which features contribute most significantly to the model's predictive power. The most prominent risk factors for ADHD prediction in Bangladeshi children were found using feature importance analysis. Features are arranged according to how much they contribute to the model's output.Table 5.6 shows that the child's gender is the strongest predictor, with a risk-level contribution of 0.6680. This aligns with epidemiological studies indicating that boys are diagnosed with ADHD at higher rates, often due to more externalised hyperactive-impulsive symptoms [13]. The high predictive value of gender in the Bangladeshi context means that the pattern of symptoms manifestation or referral could vary considerably between the sexes as is characteristic of the world trend where the often overlooked problems of girls are not adequately recognized [3]. These are the other significant features that follow gender, which are attention diverted (0.170) and age of a child (0.141). Inattention is one of the key symptoms of ADHD, which is directly related to attention being diverted. The meaning of "Child Age" means that the level of risk awareness of ADHD is highly dependent on certain age-related developmental phases or diagnostic tests. This may imply that the effects of ADHD symptoms change with age or that symptoms become more apparent or diagnosable at certain ages. Other noteworthy and significant characteristics are:

Table 5.6: Top Correlated Features with ADHD Diagnosis

Feature	**Correlation Coefficient**	**p-value**
Child's Gender	0.6680	< 0.001
Attention Level	0.5462	< 0.001
Child's Age	0.4402	< 0.001
Distraction Level	0.3375	< 0.001
Parental Genetics	0.3227	< 0.001
Focusing Ability	0.2865	< 0.001
Excessive Talk	0.2759	< 0.001
Fidgeting	0.2671	< 0.001
Study Focus	0.2567	< 0.001
Irritation	0.2234	< 0.001
Self Control	0.2130	< 0.001
Forgetting Task	0.1780	< 0.001
Income	0.1750	< 0.001

attention_details (0.127): Highlights issues with concentrating on details and is also closely linked to inattention.

excessive_talk (0.074): This behavioral indicator is frequently linked to impulsivity and hyperactivity.

tasks_forgotten (0.044): A sign of inattention and poor executive function.

Impulsivity and challenges with behavioral regulation are directly reflected in self-control (0.042).

income (0.034): Indicates that socioeconomic considerations might be important, possibly affecting environmental stimuli, stress levels, or resource access.

The prediction is also influenced by focus_ability (0.2865), study_focus (0.2567), fidgeting (0.2671), and Parental_Genetics (0.3227), demonstrating the complex nature of ADHD, which includes not just core symptoms but also associated emotional, academic, and familial contextual factors. The feature 'both_parents' is likely to include aspects of stability or family structure. These findings have relevance since they establish the most critical environmental, behavioral and demographic predictors of ADHD risk within the unique cultural and observational context of Bangladesh. This interpretability is crucial to develop targeted screening questions and direct population health campaigns. The identification of these key factors will allow the model to not only produce accurate predictions but also provide valuable insights to parents, educators and medical professionals. The conceptual output of the machine learning model to be implemented to

predict the app is supposed to be translated into three risk categories to be displayed to the user interface: Low, Moderate and High risk. The fact that the model is based on the inference of the input of the user and not the direct outcome of this feature significance analysis, although the feature importance analysis does indicate the driving force behind the predictions. The specific number of kids in each category would vary depending on which test set instances to use and what threshold is used to design each risk level in the application. This conceptual classification does not explicitly provide users with a numerical estimate of the number of individuals in each category based on this analysis step, but attempts to provide an interpretable and practical meaning of the continuous risk score of the model to guide users to the appropriate next steps.

Chapter 6

Discussion

This study successfully designed and tested machine learning models on the early detection of ADHD risk among Bangladeshi children using non-clinical survey data that is culturally appropriate. The findings have significant implications to address the existing gaps in diagnosis and intervention in low-resource settings.

6.1 Significance of Findings

The most significant finding in the study is the excellent performance of the Random Forest model that demonstrated superiority in all of the measures evaluated. Table 5.1 of the project summarizes its outstanding Accuracy of 0.91, F1-Score of 0.89 and Area Under the curve of 0.95 Judging by the collected survey data, the near-perfect performance of the discriminatory model suggests that the model can be consistent in differentiating between kids with and without ADHD. The high performance of the random Forest was further demonstrated by the comparison with support vector machine, logistic regression and decision tree models but SVM and LR models were also found to have a strong predictive power with the AUCs of 0.88 and 0.91 respectively. Importantly, the combined confusion matrix for the best algoritham in Figure 5.10 shows that the model has been able to identify all the individual children included in the test group who are diagnosed with ADHD. Minimizing false negative is important as far as early detection of neurodevelopmental disorders such as ADHD is concerned neglecting a real case may lead to untimely care, which is associated with persistent social, professional and academic problems and an increased risk of developing co-occurring mental health conditions throughout life [11] [12]. Despite the fact that, one false positive was discovered (one time, in which the model wrongly predicted ADHD), this is generally considered

less detrimental in a screening device that is supposedly preliminary, and is meant to stimulate further professional analysis, and not to provide a final diagnosis. As indicated in the Section 5.8 of the project, the feature importance analysis provided important information on the key determinants of ADHD risk in the Bangladeshi setting. The gender of the child was determined to be the most important predictor with a risk level contribution of 0.6680. This is in line with global epidemiological trends that indicate that boys are diagnosed more and often these diagnoses are associated with their more overt hyperactive-impulsive traits [13]. This illustrates how gender-specific presentation or recognition patterns of symptoms can be influencing the diagnosis in the Bangladeshi setting and reflects the trend across all regions of which the symptoms exhibited by girls are often inattentive and are unknown [3]. Other important features such as "attention diverted" and Age of child are direct indicators of the main symptoms of ADHD which are inattention and important developmental issues respectively. Such characteristics as excessive talk, tasks forgotten, self-control, study-focus and irritation also indicate the multifaceted nature of ADHD, where emotional control, hyperactivity-impulsivity and inattention are all involved. The importance of income also means that social economic factors might also have a bearing on the symptoms presented or reported and this is in line with earlier research on social determinants of health [3]. These findings will be valuable in terms of understanding the specific behavioral and environmental attributes that are most relevant in the assessment of ADHD in this group, as well as help to confirm the predictive power of the model.

6.2 Comparison with Existing Work

The performance of the model in this study puts the model in a desirable position in the sphere of machine learning-based prediction of ADHD, particularly because of its high effectiveness with the use of non-clinical, readily available survey data. Though certain previous studies with more specific neuroimaging information, such as EEG, provide similar high accuracy rates, their applicability to resource-constrained contexts, such as Bangladesh, is not always as useful. In the case of Attallah, the ADHD-AID program also utilized very large amounts of EEG data to achieve a phenomenal 0.991 accuracy [37] and Santarrosa-Lopez et al. DETEC-ADHD had reported accuracies of 100% in children in a case study that combined clinical and EEG data [36]. Though they are useful, their world scalability is restricted since they tend to require expensive equipment and expertise [32] [33]. On the one hand, our experiment demonstrates that preliminary screening can be very effective based on easily available behavioral and demographic statistics. This is a significant advancement over studies that also utilize non-neuroimaging data. As an

example, Qin et al. [40] by demonstrating a random Forest model with AUC of 0.99 and 0.95 accuracy based on demographic and behavioral rating scale data, highlighted the utility of such features. Similarly, Maniruzzaman et al. [38] achieved 85.5% accuracy on behavioral activity data with a Random Forest model. With an AUC of 0.95 and 0.91 accuracy on culturally adapted, survey-based data — and without any complex clinical or neuroimaging assessments — our Random Forest model matches and at 0.91% accuracy, exceeds the performance of many comparable models in the table 2.1 and **??**. We note that the AUC = 0.95 should be interpreted cautiously given the small sample size. This is an immediate response to a major need of readily available tools in fields where such resources are scarce. The implications of early ADHD detection are particularly high in Bangladesh. According to the problem statement, the country has severe gaps in diagnosis and intervention because of the prevalence of stigma in society and the lack of awareness [16] [17], and a considerable lack of affordable, culturally sensitive screening tools [15] [19] [21]. This study provides a practical solution to these systemic challenges by proposing a trustworthy, machine learning-powered framework on the basis of simple, questionnaire-based data. The culturally sensitive survey that is translated into Bengali ensures the tool is relevant and comprehensible to the local people, fulfilling the crucial need of validation and research that is area-specific [22] [43]. This plan supports the growing international trend of developing scalable, low-cost screening tools that do not require expert medical institutions [32]. The effective identification of children at risk through a user-friendly platform will help parents and educators to seek professional help as soon as possible, which will foster early diagnosis and intervention. This is necessary in minimizing the adverse developmental outcomes of untreated ADHD in the long-term [7] [11]. Since the localized ML-based prediction of ADHD has not been addressed in South Asia, the contribution of this project is particularly high, and it will help take a step toward more egalitarian and culturally-aware healthcare solutions.

6.3 Significance of Interface

The conceptual application interface **SHUCHOK** can serve as a key bridge and directly solve significant challenges in the Bangladeshi setting by transforming the predictive potential of the machine learning model into practical applications that can be used by end users. Its design is given the highest priority to accessibility and cultural relevance. Since Bangladesh has a diverse linguistic context and needs to ensure a broad coverage and utilization across diverse populations, bilingual support is essential. As a way of negating the stigma linked with mental health issues, the visual design of the application has been done in kid-friendly colors and symbols to destigmatize the assessment

process, making it less threatening to caregivers and more friendly to discuss children behavior [16]. As illustrated in Figure 4.1, the user friendly Low-Moderate-High meter does not overwhelm consumers with numerical data by dividing complex predicted probabilities into risk categories that are easy to understand. This is a clear disclaimer, which restates the non-diagnostic nature of the tool and provides easy follow-up actions, helps in responsibility guide the users to appropriate expert advice. In families who may not otherwise know or can afford established diagnostic pathways, or lack specialists, this approach becomes highly more aware and provides a practical starting point [15] [19] [20]. The accessibility and ease of use of the interface empower communities in identifying potential symptoms of ADHD at an early age and referring children to qualified healthcare providers, thus enhancing the developmental outcomes of children. The interface design is promising, but its ultimate usefulness is inseparably connected with the generalizability and resilience of the underlying machine learning model, which is prone to the following constraints.

6.4 Limitations

Despite the promising results, it is essential to note that there are several constraints that influence the fair performance of the model and its extrapolation. To begin with, the sample is very small, comprising only 110 people. This is illustrated in Figure 6.1:

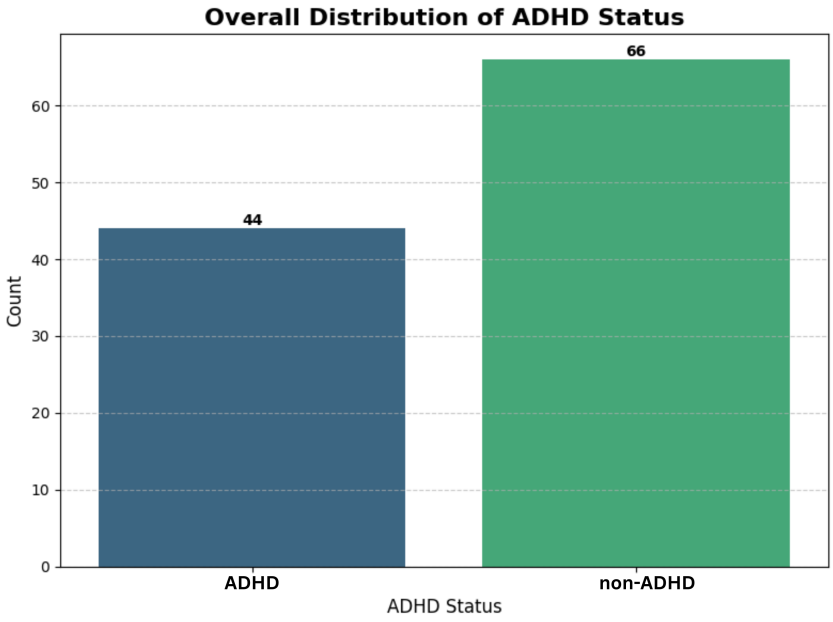

Figure 6.1: ADHD Distribution

The small sample size can restrict the external applicability of the results and may lead

to overfitting, where-by models can work well on the training data, but are not expected to work well on the hidden, larger population, as observed in the literature [23] [24] [25]. Compare to studies such as the Garcia-Argibay et al. [42] which used Sweden's nationwide registration data of 238,696 individuals, our sample is very small. A larger more diverse dataset would enhance the strength and external validity of our model which is required to be used more broadly in Bangladesh. One of the notable findings in this research is the remarkably good performance of the Random Forest model, particularly the AUC of 0.95, and the accuracy of 0.91%. Due to the relatively limited size of the dataset (N=110), these measures also give warning signs of potential overfitting despite significant classification ability. In case a model is learning patterns specific to the training data, rather than general trends, this is referred to as overfitting. Due to this, the model can perform exceptionally well on the test set and not achieve the same degree of accuracy on untested real-world data. Although measures such as feature selection and cross-validation were employed to reduce this issue, overfitting remains a possibility. To ensure strength and applicability, future studies should focus on testing the model on larger and independent datasets and external validation. Second, the subgroup with an ADHD diagnosis had a skewed distribution of males, with only 4 females, which was quite an unusual gender ratio. This is represented graphically in Figure 6.2:

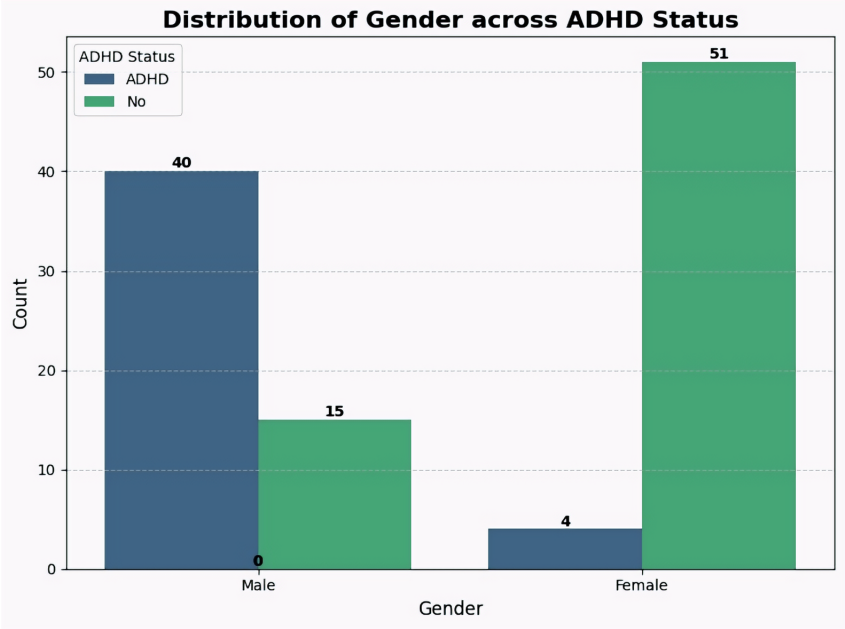

Figure 6.2: Gender Distribution by ADHD Status

Even though the 10:1 male to female ratio in the ADHD group might reflect a global trend that males are more often diagnosed due to more externalizing and hyperactivity-impulsive symptoms [13], it raises grave doubts about the fair prediction of our model on female children. Child Gender was the strongest significant predictor (risk level contribution 0.6680) which can be suggestive of diagnostic biases and not necessarily prevalence

or the diversity of symptom presentations among females. Girls internalized inattention symptoms are usually underestimated [3] [20] [27], possibly making them underrepresented in clinically-diagnosed samples. This gender imbalance may lead to a less sensitive model to ADHD presentation in girls; a problem that must be put into careful consideration and additional studies with more balanced datasets are required. Despite being promising, the impressive performance of the model (AUC 0.95, 91% accuracy) must be interpreted in the context of these constraints of the data. Although this is a positive comparison to other data-based studies on behavior, such as the 85.5 percent accuracy of Maniruzzaman et al. [38] or Qin et al.'s 95% accuracy and 0.99 AUC [40], The small, imbalanced sample can limit the generalization and pose the risk of overestimation of the practical effectiveness [45]. The significance of the observed features, in particular, the attention deflection, fidgeting, studying concentration, and irritation, is consistent with the past behavioral studies [40], Bigger, more geographically representative and gender-balanced samples are required to test its stability and predictive strength with the heterogeneous population of Bangladesh. Finally, survey-based data may have recollection and social desirability bias despite its convenience in terms of accessibility [28] [29]. The dichotomous definition of ADHD, therefore, simplifies a complex neurodevelopmental spectrum and thus disregards the variance in symptom expression and severity, critical to a holistic picture of the disorder [30] [31]. Subsequent studies will attempt to address these limitations by collecting additional information and more advanced methods of categorization.

6.5 Practical Implications in the Bangladeshi Settings

The proposed machine learning-based ADHD screening method has significant practical implications, particularly when resources are limited such as in Bangladesh. Firstly, the system provides a less expensive and readily accessible alternative to traditional diagnostic methods which often require trained health care providers and expensive assessments. Due to the limited availability of mental health services in rural and neglected communities, it is particularly advantageous. Second, the tool can be used by parents, educators, and other caregivers to evaluate children at an early age, which enables them to identify any ADHD symptoms in time. Early detection is needed to prevent behavioral, social and academic problems that will occur over time. Thirdly, the bilingual interface of **SHUCHOK** (English-Bengali) enhances inclusiveness and usability by ensuring that individuals with various language backgrounds can operate the system effectively.

Finally, the strategy can also reduce the stigma surrounding mental health problems and encourage families to consult professional help when it is needed by raising awareness on

ADHD symptoms and providing a readily available screening tool.

Chapter 7

Conclusion

This project demonstrates clearly how machine learning can be applied to address a significant public health problem in Bangladesh. We aimed to develop a data-driven, culturally sensitive and cost-effective method of diagnosing the ADHD symptoms among Bangladeshi children. The findings are a strong indication of the feasibility and great impact of this new approach.

We successfully developed trustworthy machine learning models based on survey-based, non-clinical data. The most successful model was the Random Forest with an outstanding Accuracy of 91%, F1-Score of 0.89 and an Area Under the Curve of 0.95. In a preliminary screening tool, such a high sensitivity is essential since a missed diagnosis can lead to late treatment and severe long-term outcomes, including social, professional, and academic difficulties, and the increased likelihood of comorbid mental health conditions.

Our extensive feature significance analysis has uncovered the most significant predictors of ADHD in the Bangladeshi setting. It was found that the most important factor was the Gender of the child closely followed by attention diverted, age of the child, fidgeting, study focus and irritation. Besides the fact that these results are in line with global interpretations of the symptomatology of ADHD, they help to understand the effects of specific local factors and the potential gender-specific bias in diagnoses, since often the symptoms of girls remain unnoticed. The culturally specific survey in Bengali was able to capture these contextually relevant behavioral and demographic pieces of data.

Also, the conceptual user interface, SHUCHOK, which transforms complex machine learning predictions into a bilingual, user-friendly tool to be used by parents, educators, and guardians, adds significantly to the impact of the project. Its thoughtfully developed design directly tackles the barriers of lack of knowledge, stigmatization of screening, and the crippling lack of affordable screening services in Bangladesh. SHUCHOK is an es-

sential connection to make communities request professional assessment and intervention promptly by providing a clear Low/Moderate/High risk assessment along with the corresponding disclaimers and relevant next steps. Conclusively, this effort is a major leap in enhancing the early detection and access to mental health fairly in low resource areas like Bangladesh. Our machine learning solution combined with a convenient and easy-to-use interface is a powerful, scalable, and cost-effective solution regardless of the inherent limitations such as sample size and gender imbalance. SHUCHOK can fill the existing gap of diagnosis and give more positive developmental outcomes to children with ADHD.

Chapter 8

Future Work

Several essential guidelines towards future research and enhancement are noted to enhance the soundness, generalizability, and usefulness of the ADHD prediction system based on the promising framework of this project.

- Growing and Diversifying the Dataset: The most important priority is to enlarge the existing dataset, which is currently 110 participants. This extension requires the collection of data on a much bigger, more diverse and geographically representative sample that encompasses rural and urban regions of Bangladesh. Notably, to decrease existing biases in the current diagnosis and ensure the fair prognosis of female children, whose interiorized symptoms are often overlooked, it is necessary to deliberately approach the balance between the genders in the ADHD-diagnosed sample. With a much bigger dataset, the model will be statistically more powerful, less prone to overfitting, and more robust in generalizing to the environment, making it possible to conduct more in-depth analyses of environmental and demographic variables.
- Building a Full-fledged Mobile Application: A fully functional and safe mobile app needs to be developed on the basis of the conceptual SHUCHOK interface which is being prototyped now in Canva. During this development process, the best performing Random Forest model will be easily implemented to the backend of the app, ensuring that data can be entered by users, processed properly, and the risk output clearly understood. The application must be adapted to popular mobile platforms (Android and iOS) to ensure the widest accessibility among the different Bangladeshi population.
- Rigorous Usability Testing and Clinical Validation: The functional mobile application will have to undergo rigorous usability testing with its target end-users, the

Bangladeshi parents, teachers and guardians, before any larger scale rollout. This process will be critical in obtaining feedback on the general user experience of the app, cultural sensitivity, clarity, and intuitiveness. Simultaneously, it needs to be followed by a comprehensive clinical validation study. This involves comparing risk assessment by the app to official, gold-standard clinical diagnosis by trained professionals on a different cohort. To prove the validity of the app in further healthcare recommendations, it is essential to check its realistic efficacy and reliability as a preliminary screening tool.

- Exploring Advanced Machine Learning Models: With a larger dataset, future research should examine more advanced machine learning models such as deep learning models (e.g., neural networks). These advanced models are able to better detect complex, non-linear patterns in larger datasets. At the same time, to keep openness and ensure clinical implementation of these more complex models, continuous research of interpretable AI techniques would be necessary.
- Longitudinal Studies and Broader Outcomes: Future work may include initiating longitudinal studies to track the progression of the children identified by the application as at-risk. It would then be possible to carry out an extensive evaluation of the long-term impacts of early detection with the help of the app on the outcome of interventions, academic performance, and overall living standard. Such studies would provide concrete evidence on the ultimate value of the system in enhancing the child development in the whole of Bangladesh.

Bibliography

[1] A. Koutsoklenis and J. Honkasilta, "ADHD in the DSM-5-TR: What has changed and what has not," *Frontiers in Psychiatry*, vol. 13, Jan. 2023.

[2] A. R. Hidalgo-Muñoz, D. A. Vicente, A. García-Pérez, and C. Tabernero, "Application of neurotechnology in students with ADHD: An umbrella review," *Comunicar*, vol. 31, Apr. 2023.

[3] F. Nazmin, A. Roy, T. Bushra, I. J. Retina, K. H. Arnab, and F. Ferdaus, "Exploring the prevalence and social determinants of ADHD and comorbidities among urban school-aged children in bangladesh," Dec. 2024.

[4] S. R. Pliszka, "Attention-deficit hyperactivity disorder across the lifespan," *FOCUS The Journal of Lifelong Learning in Psychiatry*, vol. 14, p. 46, Jan. 2016.

[5] A. P. Association, *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Association, 5th ed., 2013.

[6] A. Pericak, "Diagnosing attention-deficit hyperactivity disorder using the DSM-5 criteria," *The Journal for Nurse Practitioners*, vol. 11, p. 274, Feb. 2015.

[7] V. Tourjman *et al.*, "Psychosocial interventions for attention deficit/hyperactivity disorder: A systematic review and meta-analysis by the CADDRA guidelines work GROUP," *Brain Sciences*, vol. 12, p. 1023, Aug. 2022.

[8] R. Schachar, "Fifty years of executive control research in attention-deficit/hyperactivity disorder: What we have learned and still need to know," *Neuroscience & Biobehavioral Reviews*, vol. 155, p. 105461, Nov. 2023.

[9] C. J. Vaidya and M. Stollstorff, "Cognitive neuroscience of attention deficit hyperactivity disorder: Current status and working hypotheses," *Developmental Disabilities Research Reviews*, vol. 14, p. 261, Jan. 2008.

[10] K. Rubia, "Cognitive neuroscience of attention deficit hyperactivity disorder (ADHD) and its clinical translation," *Frontiers in Human Neuroscience*, vol. 12, p. 100, Mar. 2018.

[11] M. Shaw, H. Caci, P. Hodgkins, J. Kahle, N. Callamaras, and A. G. Woods, "Review of studies of ADHD: Long-term outcomes with and without treatment," *European Psychiatry*, vol. 26, p. 579, Mar. 2011.

[12] M. Shaw *et al.*, "A systematic review and analysis of long-term outcomes in attention deficit hyperactivity disorder: effects of treatment and non-treatment," *BMC Medicine*, vol. 10, Sept. 2012.

[13] G. Ayano, S. Demelash, Y. Gizachew, L. Tsegay, and R. Alati, "The global prevalence of attention deficit hyperactivity disorder in children and adolescents: An umbrella review of meta-analyses," *Journal of Affective Disorders*, vol. 339, p. 860, July 2023.

[14] N. Salari *et al.*, "The global prevalence of ADHD in children and adolescents: a systematic review and meta-analysis," *The Italian Journal of Pediatrics*, vol. 49, Apr. 2023.

[15] G. Olatunji *et al.*, "Navigating unique challenges and advancing equitable care for children with ADHD in africa: a review," *Annals of Medicine and Surgery*, vol. 85, p. 4939, Aug. 2023.

[16] R. Afreen *et al.*, "Enhancing mental health literacy and care through community-driven solutions in rural bangladesh," *Frontiers in Global Women's Health*, vol. 5, Dec. 2024.

[17] M. O. Faruk *et al.*, "Mental illness stigma in bangladesh: Findings from a cross-sectional survey," *Cambridge Prisms Global Mental Health*, vol. 10, Jan. 2023.

[18] S. Seppä, A. Halt, T. Nordström, and T. Hurtig, "Effects of symptoms of attention-deficit/hyperactivity disorder (ADHD) and oppositional defiant disorder (ODD) on academic performance and educational attainment," *Child Psychiatry & Human Development*, vol. 56, p. 696, Sept. 2023.

[19] M. Cottin *et al.*, "Digital use of standardised assessment tools for children and adolescents: can available paper-based questionnaires be used free of charge in electronic format?," *BMC Psychiatry*, vol. 22, June 2022.

[20] M. Gore and J. Morgan, "Children with attention deficit hyperactivity disorder in india: Strengthening diagnosis, support, training and research," *Journal of Indian Association for Child and Adolescent Mental Health*, vol. 21, p. 17, Feb. 2025.

[21] A. Javid, M. Ahmed, and E. Sikandar, "Breaking ground on ADHD diagnosis in pakistan," *Journal of the Pakistan Medical Association*, vol. 74, p. 613, Feb. 2024.

[22] P. Fourie and J. Armstrong, "Challenges and opportunities in attention deficit hyperactivity disorder in developing countries," *Medical Research Archives*, vol. 12, Jan. 2024.

[23] D. M. Thomas *et al.*, "Machine learning modeling practices to support the principles of AI and ethics in nutrition research," *Nutrition and Diabetes*, vol. 12, Dec. 2022.

[24] M. Mengi and D. Malhotra, "Artificial intelligence based techniques for the detection of socio-behavioral disorders: A systematic review," *Archives of Computational Methods in Engineering*, vol. 29, p. 2811, Nov. 2021.

[25] P. Charilaou and R. Battat, "Machine learning models and over-fitting considerations," *World Journal of Gastroenterology*, vol. 28, p. 605, Jan. 2022.

[26] M. Miller, "Challenges and opportunities for the early identification of ADHD across multiple levels of analysis – a reflection on tobarra-sanchez et al. (2022)," *JCPP Advances*, vol. 2, Aug. 2022.

[27] P. K. Goh, A. R. Elkins, P. S. Bansal, A. G. Eng, and M. M. Martel, "Data-driven methods for predicting ADHD diagnosis and related impairment: The potential of a machine learning approach," *Research on Child and Adolescent Psychopathology*, vol. 51, p. 679, Jan. 2023.

[28] J. Dang, K. M. King, and M. Inzlicht, "Why are self-report and behavioral measures weakly correlated?," *Trends in Cognitive Sciences*, vol. 24, p. 267, Feb. 2020.

[29] K. Arrow *et al.*, "Evaluating the use of online self-report questionnaires as clinically valid mental health monitoring tools in the clinical whitespace," *Psychiatric Quarterly*, vol. 94, p. 221, May 2023.

[30] Y. Wang *et al.*, "The neural and genetic underpinnings of different developmental trajectories of attention-deficit/hyperactivity symptoms in children and adolescents," *BMC Medicine*, vol. 22, June 2024.

[31] H. Elahi, A. Iosif, P. Mukherjee, S. P. Hinshaw, and J. B. Schweitzer, "Using hot and cool measures to phenotype and predict functional outcomes across dimensions of ADHD and typical development in adolescents," *Research on Child and Adolescent Psychopathology*, vol. 52, p. 579, Dec. 2023.

[32] M. A. Mooney *et al.*, "Prediction of attention-deficit/hyperactivity disorder diagnosis using brief, low-cost clinical measures: A competitive model evaluation," *Clinical Psychological Science*, vol. 11, p. 458, Dec. 2022.

[33] M. M. Rahman, "Enhancing ADHD prediction in adolescents through fitbit-derived wearable data," *medRxiv*, Sept. 2024.

[34] M. Cao, E. Martin, and X. Li, "Machine learning in attention-deficit/hyperactivity disorder: new approaches toward understanding the neural mechanisms," *Translational Psychiatry*, vol. 13, p. 236, July 2023.

[35] C. Q. López, V. D. G. Vera, and M. P. Quintero, "Diagnosis of ADHD in children with EEG and machine learning: Systematic review and meta-analysis," *Clínica y Salud*, vol. 36, p. 109, July 2025.

[36] I. Santarrosa-López *et al.*, "DETEC-ADHD: A data-driven web app for early ADHD detection using machine learning and electroencephalography," *Big Data and Cognitive Computing*, vol. 9, p. 3, Dec. 2024.

[37] O. Attallah, "ADHD-AID: Aiding tool for detecting children's attention deficit hyperactivity disorder via EEG-based multi-resolution analysis and feature selection," *Biomimetics*, vol. 9, p. 188, Mar. 2024.

[38] M. Maniruzzaman, J. Shin, and M. A. M. Hasan, "Predicting children with ADHD using behavioral activity: A machine learning analysis," *Applied Sciences*, vol. 12, p. 2737, Mar. 2022.

[39] O. Slobodin, I. Yahav, and I. Berger, "A machine-based prediction model of ADHD using CPT data," *Frontiers in Human Neuroscience*, vol. 14, Sept. 2020.

[40] H. Qin *et al.*, "Interpretable machine learning approaches for children's ADHD detection using clinical assessment data: an online web application deployment," *BMC Psychiatry*, vol. 25, Feb. 2025.

[41] N. de Lacy and M. J. Ramshaw, "Selectively predicting the onset of ADHD, oppositional defiant disorder, and conduct disorder in early adolescence with high accuracy," *Frontiers in Psychiatry*, vol. 14, p. 1280326, Dec. 2023.

[42] M. Garcia-Argibay, Y. Zhang-James, S. Cortese, P. Lichtenstein, H. Larsson, and S. V. Faraone, "Predicting childhood and adolescent attention-deficit/hyperactivity disorder onset: a nationwide deep learning approach," *Molecular Psychiatry*, vol. 28, p. 1232, Dec. 2022.

[43] Y. Zhang-James, A. Razavi, M. Hoogman, B. Franke, and S. V. Faraone, "Machine learning and MRI-based diagnostic models for ADHD: are we there yet?," *medRxiv*, Oct. 2020.

[44] L. Ter-Minassian *et al.*, "Assessing machine learning for fair prediction of ADHD in school pupils using a retrospective cohort study of linked education and healthcare data," *BMJ Open*, vol. 12, Dec. 2022.

[45] A. S. Morrow, A. D. C. Vega, X. Zhao, and M. M. Liriano, "Leveraging machine learning to identify predictors of receiving psychosocial treatment for attention deficit/hyperactivity disorder," *Administration and Policy in Mental Health and Mental Health Services Research*, vol. 47, p. 680, May 2020.

[46] Y. S. Liu *et al.*, "Early identification of children with attention-deficit/hyperactivity disorder (ADHD)," *PLOS Digital Health*, vol. 3, Nov. 2024.

[47] A. Weigard *et al.*, "Generalizable prediction of childhood ADHD symptoms from neurocognitive testing and youth characteristics," *Translational Psychiatry*, vol. 13, p. 225, June 2023.

[48] M. A. Mooney *et al.*, "Prediction of ADHD diagnosis using brief, low-cost, clinical measures: a competitive model evaluation," *medRxiv*, Dec. 2021.

[49] T. Chen, G. Antoniou, M. Adamou, I. Tachmazidis, and P. Su, "Automatic diagnosis of attention deficit hyperactivity disorder using machine learning," *Applied Artificial Intelligence*, vol. 35, p. 657, June 2021.

[50] N. Alsharif, M. H. Al-Adhaileh, and M. Al-Yaari, "Accurate identification of attention-deficit/hyperactivity disorder using machine learning approaches," 2024.

[51] N. Alsharif, M. H. Al-Adhaileh, S. N. Alsubari, and M. Al-Yaari, "ADHD diagnosis using text features and predictive machine learning and deep learning algorithms," vol. 3, Jan. 2024.

[52] K. Su, B. Kim, J. Lee, and H. K. Yoo, "Prediction of attention deficit hyperactivity disorder using the comprehensive attention test: a large-scale machine learning approach," *Frontiers in Psychiatry*, vol. 16, May 2025.

[53] S. García-Ponsoda, A. Maté, and J. Trujillo, "Refining ADHD diagnosis with EEG: The impact of preprocessing and temporal segmentation on classification accuracy," *arXiv*, 2024.

[54] T. S. Emser, B. Johnston, J. D. Steele, J. J. S. Kooij, L. B. Thorell, and H. Christiansen, "Assessing ADHD symptoms in children and adults: evaluating the role of objective measures," *Behavioral and Brain Functions*, vol. 14, p. 11, May 2018.

[55] Python Software Foundation, *Python Language Reference, version 3.11.* Python Software Foundation, 2024.

[56] Google Research, "Google Colaboratory." `https://colab.research.google.com`, 2024.

[57] W. McKinney, "Data structures for statistical computing in Python," in *Proceedings of the 9th Python in Science Conference (SciPy 2010)* (S. van der Walt and J. Millman, eds.), pp. 56–61, 2010.

[58] The pandas development team, "pandas-dev/pandas: pandas," 2024.

[59] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. Fernández del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.

[60] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson,

C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[61] S. Seabold and J. Perktold, "statsmodels: Econometric and statistical modeling with Python," in *Proceedings of the 9th Python in Science Conference (SciPy 2010)*, pp. 92–96, 2010.

[62] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[63] M. L. Waskom, "seaborn: Statistical data visualization," *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021.

[64] Joblib Development Team, "Joblib: running Python functions as pipeline jobs," 2024.

[65] H. Marquez Von Ralph Dane *et al.*, "Machine-learning-based investigation on classifying binary and multiclass behavior outcomes of children with PIMD/SMID," *arXiv*, Feb. 2022.

[66] B. L. Ortiz *et al.*, "Data preprocessing techniques for AI and machine learning readiness: Scoping review of wearable sensor data in cancer care," *JMIR mhealth and uhealth*, vol. 12, Aug. 2024.

[67] G. Erol, B. Uzbaş, C. Yücelbaş, and Ş. Yücelbaş, "Analyzing the effect of data preprocessing techniques using machine learning algorithms on the diagnosis of COVID-19," *Concurrency and Computation Practice and Experience*, vol. 34, Oct. 2022.

[68] F. Sogandi, "Identifying diseases symptoms and general rules using supervised and unsupervised machine learning," *Scientific Reports*, vol. 14, Aug. 2024.

[69] M. M. Ahsan, M. A. P. Mahmud, P. Saha, K. D. Gupta, and Z. Siddique, "Effect of data scaling methods on machine learning algorithms and model performance," *Technologies*, vol. 9, p. 52, July 2021.

[70] M. A. Cruz, O. A. C. Hurtado, and E. C. Madariaga, "Application of machine learning in learning problems and disorders: A systematic review," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 12, 2023.

[71] L. Caselles-Pina, P. S. del Amo, D. Aguado, J. López-Castromán, J. d. D. Sanjuán-Antúnez, and D. Delgado-Gómez, "A data science approach to optimize ADHD assessment with the BRIEF-2 questionnaire," *Translational Neuroscience*, vol. 15, p. 20220349, Jan. 2024.

[72] D. Andrikopoulos, G. Vassiliou, P. Fatouros, C. Tsirmpas, A. Pehlivanidis, and C. Papageorgiou, "Machine learning-enabled detection of attention-deficit/hyperactivity

disorder with multimodal physiological data: a case-control study," *BMC Psychiatry*, vol. 24, Aug. 2024.

[73] J. Sadaiyandi, P. Arumugam, A. K. Sangaiah, and C. Zhang, "Stratified sampling-based deep learning approach to increase prediction accuracy of unbalanced dataset," *Electronics*, vol. 12, p. 4423, Oct. 2023.

[74] V. W. de Vargas, J. Aranda, R. dos Santos Costa, P. Pereira, and J. L. V. Barbosa, "Imbalanced data preprocessing techniques for machine learning: a systematic mapping study," *Knowledge and Information Systems*, vol. 65, p. 31, Nov. 2022.

[75] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[76] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[77] J. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *Advances in Large Margin Classifiers*, pp. 61–74, MIT Press, 1999.