AI-Driven Early Age ADHD Prediction and Analysis: A Novel Approach Integrating Machine Learning, Explainable AI, LLMs, and a Dialogflow-Based Virtual Therapist Chatbot (ComfortChat)

Senior Design Project

Md. Adham Wahid ID# 2111177042
Fatema Afsan Ema ID# 2022281642
Istiak Ahasan ID# 2012082642

Faculty Advisor:
Dr. Sifat Momen
Associate Professor
ECE Department

Department of Electrical and Computer Engineering
North South University

Spring 2025

LETTER OF TRANSMITTAL

April, 2025

To
Dr. Mohammad Abdul Matin
Chairman
Department of Electrical and Computer Engineering
North South University, Dhaka

Subject: Submission of Capstone Project Report titled -
"AI-Driven ADHD Prediction and Analysis at Early Age: A Novel Approach Integrating Machine Learning, Explainable AI, LLMs, and Dialogflow with a Virtual Therapist Chatbot (ComfortChat)"

Respected Sir,

We are pleased to submit our Capstone Project Report titled *"AI-Driven ADHD Prediction and Analysis at Early Age: A Novel Approach Integrating Machine Learning, Explainable AI, LLMs, and Dialogflow with a Virtual Therapist Chatbot (ComfortChat)"* as part of the requirement for completing our Bachelor of Science in Electrical and Computer Engineering.

This project reflects our effort to develop an AI-powered solution for early ADHD detection and therapeutic support. Supervised by Dr. Sifat Momen, this work incorporates machine learning, LLMs, and Dialogflow to propose a practical, real-world system.

We will be highly obliged if you kindly receive this report and provide your valuable judgment. Moreover, thank you and the department for your continuous support and guidance.

Sincerely,

Md. Adham Wahid ID# 2111177042
Fatema Afsan Ema ID# 2022281642
Istiak Ahasan ID# 2012082642

Department of Electrical and Computer Engineering
North South University

APPROVAL

This is to certify that Md. Adham Wahid (ID #2111177042), Fatema Afsan Ema (ID #2022281642), and Istiak Ahasan (ID #2012082642) of the Department of Electrical and Computer Engineering at North South University have successfully completed the Senior Design Project titled "*AI-Driven ADHD Prediction and Analysis at Early Age: A Novel Approach Integrating Machine Learning, Explainable AI, LLMs, and Dialogflow with a Virtual Therapist Chatbot (ComfortChat)*" under the supervision of Dr. Sifat Momen. This work was carried out in partial fulfillment of the requirements for the degree of Bachelor of Science in Electrical and Computer Engineering and has been accepted as satisfactory.

Supervisor's Signature: ______________________________

Dr. Sifat Momen
Associate Professor
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh

Chairman's Signature: ______________________________

Dr. Mohammad Abdul Matin
Professor & Chair
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh

DECLARATION

This is to declare that this project is our original work. No part of this work has been submitted elsewhere, partially or fully, for the award of any other degree or diploma. All project-related information will remain confidential and shall not be disclosed without the formal consent of the project supervisor. Relevant previous works presented in this report have been properly acknowledged and cited. The plagiarism policy, as stated by the supervisor, has been maintained.

Students' Names & Signatures

- Md. Adham Wahid

- Fatema Afsan Ema

- Istiak Ahasan

ACKNOWLEDGEMENTS

The authors would like to express their heartfelt gratitude towards their project and research supervisor, Dr. Sifat Momen, Associate Professor, Department of Electrical and Computer Engineering, North South University, Bangladesh, for his invaluable support, precise guidance, and advice pertaining to the experiments, research, and theoretical studies carried out during the course of the current project, and also in the preparation of the current report.

Furthermore, the authors would like to thank the Department of Electrical and Computer Engineering, North South University, Bangladesh, for facilitating the research. The authors would also like to thank their loved ones for their countless sacrifices and continual support.

ABSTRACT

A developmental illness that may manifest in early childhood is Attention Deficit Hyperactivity Disorder (ADHD). It may lead to poor self-esteem and social function in children if it's not considered at an early age. Early detection and diagnosis of ADHD results in early interventions that improve social development, academic performance, and treatment outcomes among children. This is the first integrated framework that not only predicts ADHD in early childhood using Machine learning and Natural language processing techniques but also offers a Dialogflow-based therapeutic chatbot as a primary intervention an approach not addressed in existing literature. A total of fourteen Machine Learning and five BERT-based transformer algorithms with optimized hyperparameters have been applied to classify ADHD-positive and ADHD-free children identifying the key factors. The Stacking ensemble model came out with the best-performing metrics with 94.27% accuracy, 87.00% precision, 82.07% recall and 84.30% f1 score, and AUC of 0.97 that achieves state-of-the-art performance on CAHMI survey and dataset. While DistilBERT outshined among other BERT-based transformer models with 93.45% accuracy, 19.25s runtime, and AUC of 0.968 highlighting the natural language processing. LIME and SHAP have added enhanced transparency and dynamic decision-making interpretability. Unlike prior studies that focus solely on prediction, our model emphasizes both early diagnosis and immediate therapeutic support, bridging a critical gap between detection and intervention. Moreover, domain experts raw feedback and generalization through unseen datasetscontribute significantly to the novelty and practical relevance of the proposed system. This study presents a comprehensive statistical analysis through a well-structured machine learning and transformer pipeline, incorporating detailed visualizations alongside textual explanations.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

The neurodevelopmental disorder known as Attention-Deficit/Hyperactivity Disorder (ADHD) [1] is typified by age-inappropriate, impaired inattention and/or hyperactivity/impulsivity. Theoretical perspectives on distinct brain dysfunctions have given way to more intricate models that take into account the variety of ADHD's clinical manifestations as a result of neurobiological research conducted in recent decades [2] . Nowadays, ADHD can be considered a common mental disorder among people. It is noticeable among people of different ages, genders, races, and family structures [3]. Often it impacts day-to-day life [4], including engagement in the community, relationships with family and friends, and performance at work or school.

As it is usual to find people with ADHD, detecting the behaviors at an early age [5] helps to solve the issues with proper care. Adolescents with ADHD who receive their diagnoses later than those who receive them feel more alone and have worse self-esteem [6]. Emphasizing early detection [5] of ADHD has a limitless impact on the diagnosis and recovery system.

The diagnostic criteria for ADHD include symptoms of inattention and/or hyperactivity/impulsivity, functional impairment across several contexts, and symptoms starting prior to the age of 12 years [7]. Early diagnosis and treatments are absolutely vital for reducing the long-term effects of ADHD on social contact, mental health, and academic achievement. The prevalence of ADHD diagnoses among children in the United States has progressively risen, from an estimated 6–8% in 2000 to approximately 9–10% in 2018 [8][9][10][11]. Implications of ADHD prediction at an early age have several benefits. Having access to an early diagnosis of ADHD enables an individual to maximize the benefits of treatments such as behavioral therapy, medication, and special educational help [12]. It is, therefore, vital to make ADHD predictions as early as possible if we want children to enjoy a better quality of life and achieve enduring success later on.

Children have limited access to mental health resources for their psychiatric conditions, particularly for neurodevelopmental disorders like ADHD. Many children remain undiagnosed or receive delayed treatments due to a lack of awareness, stigma, or lack of adequate specialized centers of healthcare for their condition [13].

1.2 Purpose and Goal of the Project

This study delves into the earlier detection and intervention of ADHD, particularly in children, as an early diagnosis helps improve long-term prognosis outcomes. Unlike traditional methods that depend greatly on personal judgments, the proposed system is based on artificial intelligence, which is more objective, clear, and affordable. In particular, it utilizes behavioral, environmental, and socio-emotional markers of ADHD and applies machine learning, Explainable AI, and Natural Language Processing (NLP) to improve understanding and usability features. One of the major tasks of the study involves the design of a virtual therapy chatbot in Dialogflow which intends to customize the assistance by guiding the intervention, teaching, and caring access for children, parents, caregivers, and others. This chatbot helps children and their caregivers check and support self-evaluation, self-help, and self-therapy at early ages while helping reduce the stigma and clinic appointment overuse.

The proposed study addresses major gaps and opens up opportunities for more comprehensive, clear, and easy-to-use diagnosis and intervention protocols for ADHD.

Using machine learning and transformer techniques, the system can close gaps in the diagnosis of ADHD and develop a low-cost, high-scalability solution that is applicable in socio-culturally diverse settings.

The objectives of this study are as follows:

- **Research Question 1:** Is ADHD in children predictable through machine learning methods based on behavioral, environmental, demographical, and neurodevelopmental factors?
- **Research Question 2:** How does the integration of Explainable AI enhance the interpretability of ADHD prediction models?
- **Research Question 3:** Can large language models (LLMs) effectively predict ADHD likelihood and assist in early-stage screening?"

- **Research Question 4:** How effective is a virtual therapist chatbot in providing ADHD-related guidance and support?
- **Research Question 5:** What role do socio-environmental factors, such as adverse childhood experiences (ACEs), parental involvement, and healthcare accessibility, play in ADHD prediction?
- **Research Question 6:** To what extent do domain expert consensus and robust validation across unseen heterogeneous datasets ensure the translational viability of AI models for ADHD detection in clinical practice?

To answer these questions, this study takes a well-rounded approach by combining statistical machine learning, symptom analysis using Natural Language Processing (NLP), and explainable AI techniques. Key contributions of the proposed system and paper include:

- **ADHD Prediction Framework:** Develop an ML-based system that integrates behavioral and neurodevelopmental aspects of ADHD classification.
- **Explainability and Trust in AI:** Implemented Local Interpretable Model-Agnostic Explanations (LIME) and further employed SHAP to increase model explainability and reliability.
- **Conversational AI for Mental Health:** Has built a Dialogflow-based virtual therapy chatbot that engages in conversation with individuals to offer symptom assessment and ADHD depression-related resources.
- **Culturally Contextualized AI Application:** This study focuses on ADHD to target socio-cultural obstacles to diagnosis as well as treatments.
- **Statistical and Visualization Impacts:** Considering the target audiences, the system has emphasized statistical analysis with proper graphing and visualization techniques. The target is not only providing the message through texts but also showing the raw survey analysis through various tables, impactful diagrams, and flowcharts with labels.
- **Prototype Application:** The whole package has also been deployed into both web and mobile phone applications to make it convenient for users. The prototype application is solid and user-friendly.
- **Generalization and Validation:** The proposed system has been generalized with two different datasets from two different surveys. Also been validated through domain experts and caregivers for feedback.

1.3 Organization of the Report

It's a comprehensive and unique study with proper interpretation and clarity. In addition to the textual presentation, emphasis has also been placed on visual presentation to ensure the content is accessible and easily interpretable for healthcare professionals, patients, and other medical stakeholders.

The rest of the paper is structured into distinct sections as follows: Literature Review, Methodology, Performance Evaluation, Model Interpretability, Deployment and Integration, Project Impacts, Ethical Considerations, Discussion, Limitations, and Conclusion with Future Work.

Chapter 2

Literature Review

Attention deficit hyperactivity disorder is a neurodevelopmental disorder characterized by executive dysfunction occasioning symptoms of inattention, hyperactivity, impulsivity, and emotional dysregulation that are excessive and pervasive, impairing in multiple contexts, and developmentally inappropriate. If it can be detected in the early stage, it may assist and make improvements in the early diagnosis process system. Many comprehensive works and studies have been performed to detect ADHD using measuring scaling systems and artificial intelligence techniques. This section briefly discusses some of the novel studies in this area.

2.0.1 ADHD Diagnosis using the Behavior and Health Data through ML

Within Grazioli et al.'s[14] study, 82% accuracy in the classification of ADHD has been obtained from a Decision Tree model trained upon data reported by parents and teachers. Telemedicine is now an ubiquitous term among child neuropsychiatry, including online platforms of remote data aggregation. The study highlights the promising role of telehealth in using reliable caregiver reports and conducting computerized risk assessments for the diagnosis of ADHD. In spite of the difficulties presented by ASD symptoms in the diagnostic process, a cross-validation accuracy of 74% was nonetheless attained.

Maniruzzaman et al.[15] applied machine learning techniques to predict ADHD in children from behavioral and health-related factors in the 2018–2019 National Survey of Children's Health (NSCH). They balanced an imbalanced dataset (11.4% ADHD cases) through oversampling and undersampling techniques. Eight machine learning models were used, including RF, DT, XGB, and others using predictors such as anxiety, depression, asthma, low birth weight, family structure, and other features. The Random Forest has come out with the best performance with an accuracy of 85.5% and an AUC of 0.94. It has shown the potential of ML-based solutions for early detection of ADHD and interventions.

In the study by Das et al.[16], they implemented a web application to diagnose ADHD using a machine learning-based model. They collected a sample of 160 trials per patient and used multiple machine learning algorithms and pre-trained models like Logistic Regression, KNN, SVM, and many others. This application captures pupil size using CNN. They achieved diagnostic performance metrics with AUROC (Area Under the Receiver Operating Characteristic Curve) with a sensitivity of 0.821, reflecting the aura of the study.

Qin et al.[17] applied machine learning techniques for ADHD and subtypes classification for children from data on the ADHD-200 dataset clinical assessment scale. They also removed imbalances from the data using feature selection and preprocessing techniques. Seven machine learning models were tested, with the Random Forest (RF) model performing best in the classification of ADHD (AUC = 0.99) and the Support Vector Machine (SVM) being best for subtype classification (micro-average AUC = 0.96, accuracy = 0.85). The SHAP method was used by the study to render the models interpretable and indicated that the ADHD predictions had correlations with higher ADHD Index and lower IQ scores. The last model was deployed as a web application for real-time clinical use.

The study by Kim S et al.[18], used 3 machine learning algorithms for statistical analysis such as K-nearest neighbors (KNN), linear discriminant analysis(LDA), and random forest. Their accuracy was 93.1%,91.2%, and 93.6% respectively. To prevent overfitting, a validation dataset was added as a verification step. The AUC of the KNN, LDA, and Random forest method was an excellent level of diagnostic accuracy subsequently 0.722,0.806,0.790.

2.0.2 ADHD and Coexisting Conditions using Hybrid and Complex Models

The study by Shin et al.[19] explores the detection of ADHD in children in a unique manner, which is with coexisting autism spectrum disorder (ASD) using machine learning-based approaches. It focuses on handwriting patterns. The dataset was generated using handwritten samples from 29 Japanese children (14 with ADHD and ASD, 15 healthy). They drew zigzag and periodic lines on a pen tablet. From there, 30 statistical features were extracted and analyzed using sequential forward floating search (SFFS). Seven ML models were used, and the Random Forest classifier achieved the best accuracy of 93.1%. And other metrics shown score of recall: 90.48%; precision: 95.00%; f1-score: 92.68%; and AUC: 0.930. In the work of Duda et al.[20], machine learning algorithms recorded as high as 96.5% accuracy in differentiating between autism spectrum disorder (ASD) and ADHD based on data from the Social Responsiveness Scale (SRS). The researchers highlighted the need for simplification of diagnostic procedures since conventional examination can last more than a year, postponing early intervention. The study found that only five behavioral features were necessary for successful classification using forward feature selection and cross-validation. The authors highlighted the potential of mobile screening tools for brief, off-site risk assessments based on caregiver reports, while noting the need for further validation on larger populations.The late-diagnosed adolescents with ADHD feel alone more than the early-diagnosed adolescents, and they like themselves less compared to the early-diagnosed group.

Fink et al.[21] employed machine learning and hybrid models to diagnose ADHD from EEG signals of 121 children (61 ADHD, 60 controls) recorded at a sampling rate of 128 Hz. They employed preprocessing, dimensionality reduction using UMAP, and four supervised learning models: Logistic Regression, Random Forest, XGBoost, and CatBoost. The best accuracy of 73.71% and ROC-AUC of 0.7999 was achieved by the Random Forest model, followed by XGBoost (71.80% accuracy, ROC-AUC 0.7868). The study validated the feasibility of EEG-based machine-assisted detection of ADHD but identified shortcomings concerning noise, management of EEG artifact, and clinician-investigator settings to clinical settings generalizability.

2.0.3 Use of Neuroimaging and Advanced Data for ADHD Diagnosis

In the extensive study by Joy et al.[22] a computational method to detect ADHD from EEG signals using nonlinear entropy measures was proposed. Fuzzy entropy, log energy entropy, permutation entropy, SURE entropy, and Shannon entropy features were used to discriminate ADHD subjects from the control group. In the study, it was found that the permutation entropy provides the highest class accuracy of 99.82% with the highest sensitivity and specificity of 98.21% and 98.82%, respectively. Further, statistical comparison against the t-test revealed that Shannon entropy had a higher P-value (.001), thereby yielding poorer performance in ADHD diagnosis. The study further highlighted the evidence that significant differences in EEG responses in frontal polar (FP) and frontal (F) regions under eyes-closed conditions might help enhance ADHD classification. The above facts lend support to entropy-based EEG feature extraction as an emerging technique in automated ADHD detection.

In M. Maniruzzaman et al.[23], researchers have tried to enhance ADHD diagnosis with the help of biomarker discovery. RST data and fNIRs measurements from children have been taken where 72 are affected by ADHD and 171 are healthy children. In the research, the concern is about the age, behavior, and hemoglobin level in the prefrontal cortex of children. Six ML models have been trained using cross-validation after finding the important biomarkers. The Random Forest model performed optimally with the best accuracy of 91.67% and sensitivity of 95.92%. It has shown potential for ADHD diagnosis with an AUC of 0.955, suggesting a high potential for clinical strategies and applications.

Nash et al.[24] have provided an extensive survey of machine learning solutions to ADHD and depression diagnosis, reporting on the success of SVMs and CNNs applied to a variety of data sources, ranging from EEG to fMRI and clinical data. High accuracy levels are reported, e.g., 99.46% accuracy obtained by EEG-based models, testifying to the potential of ML for improving diagnosis accuracy. The authors also discuss the value of dense, multimodal data and mention privacy-sensitive mental health data issues. They conclude that although there is potential in ML models for faster and more mental health diagnoses, additional validation in the form of larger and balanced datasets is needed to enhance reliability and generalizability.

Garcia-Argibay et al.[25] has utilized the Swedish national registry data for creating a deep-learning model for predicting ADHD development in children and adolescents. Having the predictor set with most notably parental criminality, school performance, and comorbidities, the model performed with a balanced accuracy of 0.69 and AUC of 0.75, with the DNN outperforming the other models. Parental ADHD, criminality, male gender, and learning disabilities were the most predictive variables and reflect the multi-dimensional interplay of genetic and environmental factors. The model enables the introduction of early detection of ADHD into clinical practice if it can be applied to heterogeneous populations.

In the study by Mikolas et al.[26], 19 features have been selected to ensure the robustness and effectiveness of the proposed model. They have witnessed about 20% missing values from there they removed them for the performance. They have used the SVM classifier which gave 66.1% accuracy and 0.66 of Area under the curve (AUC) They achieved diagnostic performance metrics. The model emphasized the advanced data and dataset.

2.0.4 Strategies for Robust Modeling and Reliable Predictions

In the study by Khanna and his teammates[27], they used a machine learning-based framework that uitilizes -size dynamic objective which used for automated detection of ADHD. They used binary classification model to determine the ADHD positives. The data set was declassified and had 783 features from them 218 features were differentiators of ADHD. But the problem arose when they faced the dataset had more than 80% data missing and to estimate missing values they used cubic spline interpolation analysis. Moreover, they used SVM classifier with AUROC 0.856 $\pm$ 0.16, which shows the robustness of performance and CV (cross-validation) also. For classification, they used logistic regression, K nearest neighbor, and many other machine learning algorithms. For model evaluation, they used Inner and Outer CV and Evaluation matrix.

Kim et al.[28] applied machine learning to predict ADHD and sleep problems in children from wearable data from the Adolescent Brain Cognitive Development (ABCD) study. They trained models from features extracted from wearable technology based on circadian rhythm. They had 12,348 samples for ADHD and 39,160 for sleep disorders in the database. The top performance was done by LightGBM with an AUC of 0.798 for diagnosing ADHD (sensitivity: 0.756, specificity: 0.716) and an AUC of 0.737 for sleeping disorders (sensitivity: 0.743, specificity: 0.632). The study highlighted limitations in accuracy and model interpretability but also showed that early screening of ADHD using wearable-derived digital phenotypes is feasible.

In the work by Zhang-James and their group mates [29], they encountered missing values which could reduce their record size and create biases so they created "missing" as an ordinal category. Moreover, they used random forest, hyperparameter tuning, and feature selection to handle missing data. For predicting new SUD diagnoses The model's AUC's were from 0.65 to 0.66. The precision of the model was 54.6% which is relatively high at low sensitivity.

Table 2.1 provides a summary and comparison of the existing related works stated in this study.

Table 2.1: A comparative analysis of existing works highlighting its advantages and limitations

REF	Year	Dataset	Applied Technique	Metrics	Limitation
[14]	2023	Custom	Decision Tree using teacher and parent data via telemedicine	Accuracy: 82%, Cross-validation Accuracy: 74%	ASD symptom overlap
[15]	2022	NSCH 2018–2019	RF, XGBoost on balanced data	RF Accuracy: 85.5%, AUC: 0.94	Limited generalizability
[16]	2020	Custom	Web app using LR, KNN, SVM, CNN on pupil size	AUROC: 0.821, Sensitivity: 0.821	Requires clinical tuning
[17]	2025	ADHD-200	RF, SVM + SHAP for interpretation	RF AUC: 0.99, Accuracy: 96%; SVM AUC: 0.96, Accuracy: 85%	Small ADHD-H sample; no real-world validation
[18]	2021	Kongju Univ. Survey	KNN, LDA, RF	Accuracy: 93.1%, 91.2%, 93.6% (across models)	Limited validation set
[19]	2023	Japanese school students	CNN + Bi-LSTM on fNIRs + SFFS	Accuracy: 93.1%, Precision: 95.0%, Recall: 90.5%, F1-score: 92.7%, AUC: 0.93	Small sample size
[20]	2016	Social Responsiveness Scale	Feature selection + cross-validation	Accuracy: up to 96.5%	Small sample, limited scope
[21]	2025	EEG (Shahed Univ.)	ML + UMAP + tuning	RF Accuracy: 73.71%, AUC: 0.80; XGBoost Accuracy: 71.80%, AUC: 0.79	No clinical validation, low explainability
[22]	2022	Custom	6 ML models on hemoglobin biomarkers	Accuracy: 99.82%, Sensitivity: 98.21%, Specificity: 98.82%	Limited generalizability
[23]	2024	Reverse Stroop (fNIRs)	RF on ADHD vs TD children	Accuracy: 91.67%	Dataset validation needed
[24]	2023	Custom	SVM, CNN on EEG/fMRI/clinical notes	Best Accuracy: 99.46% (EEG-based model)	Privacy and validation concerns
[25]	2023	Swedish Registers	LR, RF, DNN on registry data	DNN AUC: 0.75, Balanced Accuracy: 0.69	Context-specific to Sweden
[26]	2022	TU Dresden medical records	SVM after 20% data removal	Accuracy: 66.1%, AUC: 0.66	Data removal impacted results
[27]	2020	Chilean study	SVM with spline interpolation + CV	AUROC: 0.856 $\pm$ 0.16	High missing data
[28]	2023	ABCD Study	ML + HyperSMURF + circadian features	AUC: 0.798, Sensitivity: 0.756, Specificity: 0.716	Class imbalance, low interpretability

Chapter 3

Methodology

3.1 Methodology

This section presents a comprehensive description of the proposed framework for detecting and analyzing Attention Deficit Hyperactivity Disorder (ADHD), as outlined in Figure 3.1. It encompasses the key components of the system, including data collection, preprocessing, machine learning-based classification, large language model (LLM) integration, explainable AI (XAI) techniques, and system deployment through a Dialogflow-powered therapeutic chatbot.

The methodology follows a logical and sequential flow, beginning with data acquisition and concluding with the deployment of the complete system into both web and mobile platforms. Several components are interconnected and function as intermediary bridges between different modules, ensuring a cohesive and modular system design.

This structured approach not only defines the backbone of the research but also establishes the essential mechanisms that drive each phase of the system. The proposed methodology is novel in its integration of both predictive analytics and real-time interactive support, making it, to the best of our knowledge, the first of its kind to offer a unified solution that combines machine learning predictions with an intelligent, chatbot-based therapeutic interface.

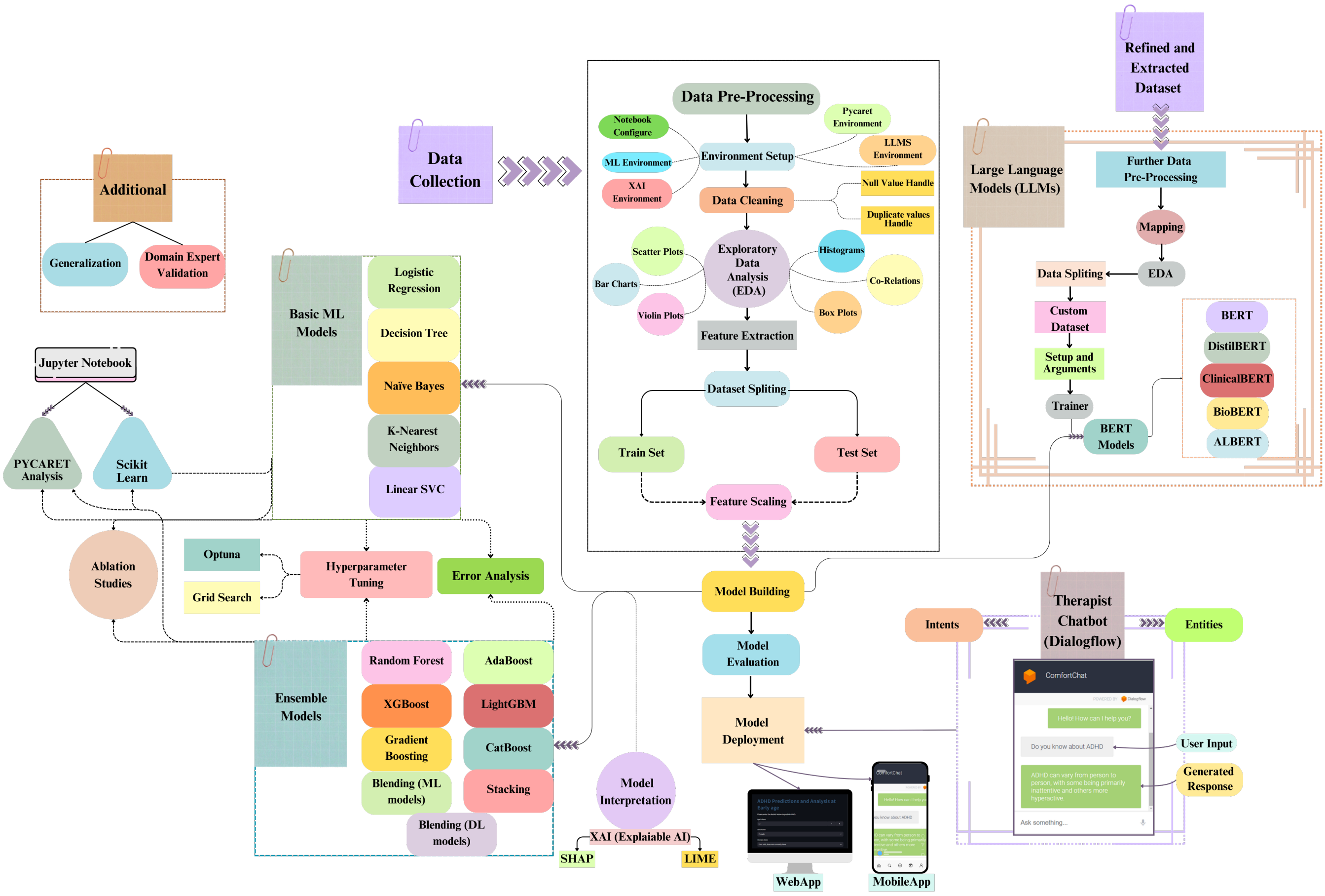

Figure 3.1: Methodology diagram of the proposed system.

3.1.1 Data collection

Data collection was conducted through a formal request to the Child and Adolescent Health Measurement Initiative (CAHMI) [30], housed at Johns Hopkins Bloomberg School of Public Health [31]. CAHMI is a leading authority in child health data, supporting initiatives such as ACEs research, the Maternal and Child Health Measurement Network, and the National Survey of Children's Health (NSCH).

For this study, we primarily used the 2021–2022 NSCH dataset, comprising 104,995 rows and 826 variables. After filtering based on literature and SPSS codebook review, a refined dataset of 73 relevant features was used for model training. To evaluate generalizability, we also employed two additional datasets: the 2022–2023 NSCH (most recent) and the 2018–2019 NSCH (historical reference).

3.1.2 Data pre-processing

Data preparation includes data preprocessing, which is any kind of processing done on raw data to prepare it for another data processing step. This study has applied and explored different types of data preprocessing techniques. They include environment setup, data cleaning, exploratory data analysis, feature extraction, dataset splitting, and feature scaling.

Environment setup

To keep track and ensure smooth transition environment setup is a must. Have used the Jupyter Notebook as the coding platform . Also set up five different environments includes the ML environment, XAI environment , Pycaret environment, LLMs environment and general configured environment. The Figure 3.2 shows the used environments for this study.

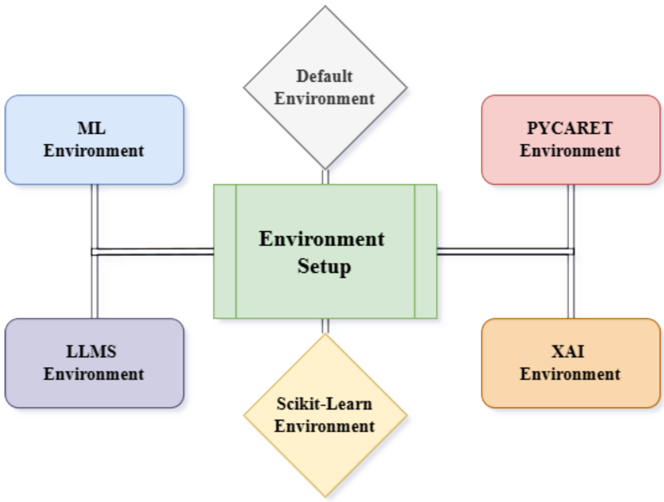

Figure 3.2: Environment setup

Data cleaning

The dataset comes with a code-book where a detailed information about the survey has been illustrated . Initially have analyzed the code-book and found out there are some specific encoded values which defines the missing values . To be more specific the values are [90,95,96,99]. As the dataset has decent amount of values , so decided to apply both null value extraction and imputation through the most frequent value . Extracted [95,96,99] holding values and imputed the values holding the key value 90. The dataset size turned into 67336 rows x 73 columns containing all the relevant features. Duplicate values has been also plucked out for ensuring unbiased result

. The dataset size been updated to 67171 rows x 72 columns removing another redundant feature . Proper cleaning of the data [32]guarantees that data is in its appropriate form depicting clearly the current patterns and relationships. Better reliability, there is an enhancement of data quality towards achieving that it is more objective and accurate. Aside from that, this attempt to clean enhances the strength of analytics from the data, we now ensure that subsequently constructed models were based on trust worthy data. This prepared data on an additional cleansing with our analysis can move into a good. a more advanced level of analysis and create good predictive models for the sake of deriving useful and relevant results. The Figure 3.3 illustrates the data cleaning process in the data pre-processing.

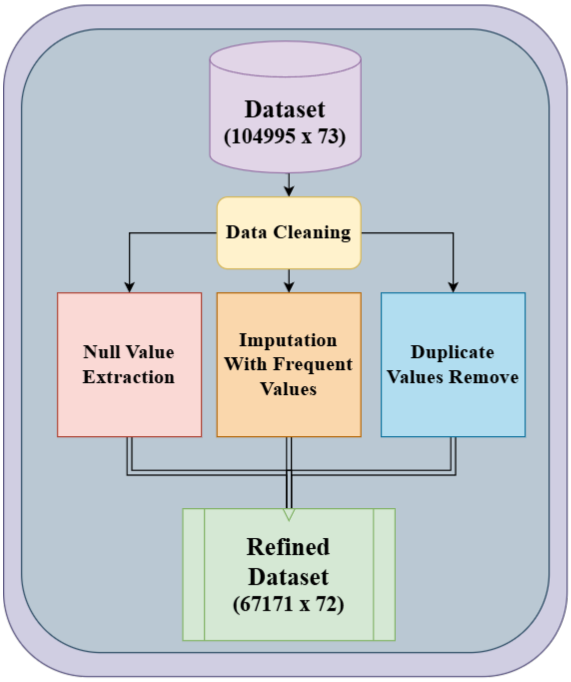

Figure 3.3: Data cleaning process

Encoding attributes

Encoding [33] is a crucial part during the data pre-processing . Fortunately, received the encoded dataset version from CAHMI [30] . The dataset has a code-book that illustrates the feature informations and relevant data.

Exploratory data analysis (EDA)

Demograpchic and Socioeconomic analysis

While analyzing and extracting the key characteristics of ADHD, the socioeconomic and demographic features play a crucial role. Simply it's a key element for the insights and proper understandings. Demographic features like age of the child, age of the mother, the family structure of the child, race and ethnicity, financial situation of the family, number of family members in the family, and parental nativity are very basic and solid attributes for the insights.

Table 3.1: Socio-Demographic Distribution (AGE and GENDER)

Category	**Group**	**Frequency (%)**
SC_AGE_YEARS (Age of child)	3-5	17,870 (26.81%)
	6-12	26,259 (39.39%)
	13-17	22,537 (33.81%)
	Total	66,666 (100.00%)
MOMAGE (Age of mother)	18-25	13,890 (20.84%)
	26-35	38,105 (57.18%)
	36-45	14,671 (22.01%)
	Total	66,666 (100.00%)
sex_2122 (Gender Distribution)	Male	34,530 (51.80%)
	Female	32,136 (48.20%)
	Total	66,666 (100.00%)

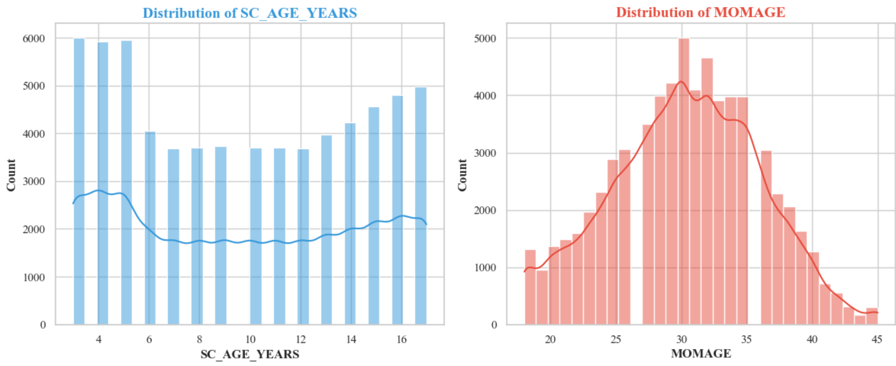

Figure 3.4: Age distribution of child and mother

From the Table 3.1 and Figure 3.4, the intuition of the ages for both the children and mothers can be retrieved. There are balanced scatter of the child age distribution, while the mother's age seem to have density in the 26-35 years age group reflecting typical childbearing age patterns. The higher proportion falls within the school-going age in the children group section. While still considerable, the percentage of older moms and young moms are comparatively smaller.

Table 3.2: Race and family structure distribution with ADHD positive cases

Category	**Group**	**Frequency (%)**
race4_2122 (Race)	White	44,952 (67.41%)
	Hispanic	9,071 (13.61%)
	Other	9,036 (13.56%)
	Black	3,607 (5.42%)
	Total	66,666 (100.00%)
ADHD Positive (by Race)	White	5,211 (11.59%)
	Hispanic	842 (9.28%)
	Other	729 (8.07%)
	Black	393 (10.90%)
famstruct5_2122 (Family Structure)	Two Parents (Married)	48,321 (72.50%)
	Single Parent	12,888 (19.34%)
	Two Parents (Unmarried)	3,480 (5.22%)
	Grandparent	1,486 (2.23%)
	Other	491 (0.71%)
	Total	66,666 (100.00%)
ADHD Positive (by Family Structure)	Two Parents (Married)	4,498 (9.31%)
	Single Parent	1,805 (14.00%)
	Two Parents (Unmarried)	413 (11.87%)
	Grandparent	247 (16.63%)
	Other	49 (9.98%)

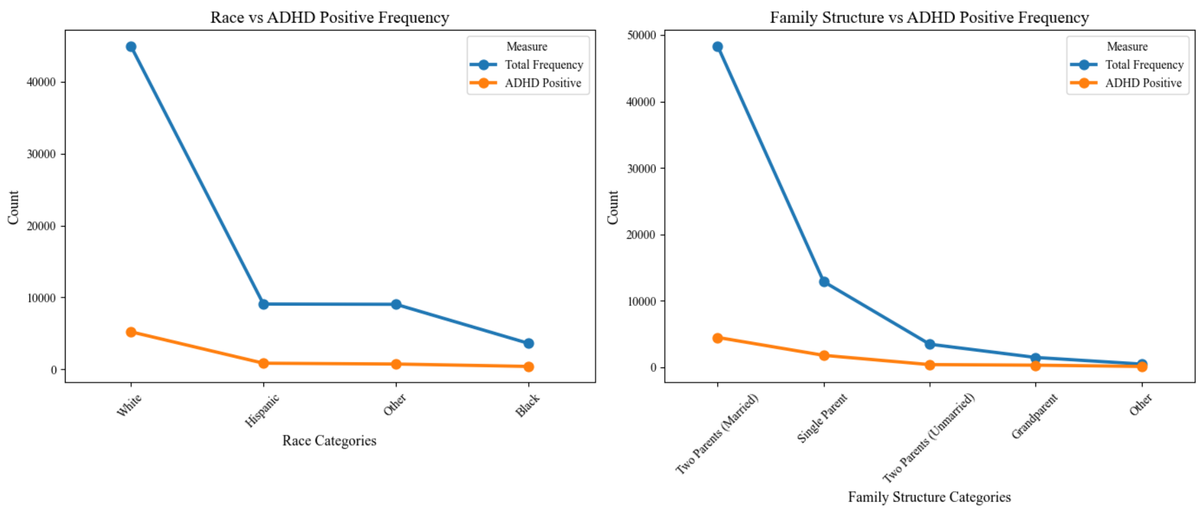

Figure 3.5: Race x Family structure distribution (ADHD)

The analysis reveals that ADHD-positive cases are most prevalent among White children and those from two-parent married households when considering absolute frequencies. However, Figure 3.5 demonstrates that prevalence rates tell a more nuanced story: while the absolute counts are lower, children in grandparent-headed households show the highest ADHD rate (16.63%), with minimal variation between this and other family structures. Complete demographic breakdowns by race and family structure are provided in Table 3.2.

Adverse Childhood Experiences (ACEs) and Social Support

Table 3.3: Overall and ADHD positive frequency adverse childhood experiences and social support

Feature	Overall	ADHD Positive	Percentage (%)
ACE2more6HH_2122 (Household Based)			
No Household ACE	45,977	3,489	7.59%
1 Household ACE	12,083	1,688	13.97%
2+ Household ACEs	8,606	1,998	23.22%
ACE1more4Com_2122 (Community Based)			
No Community ACE	59,954	5,277	8.80%
1+ Community ACEs	6,712	1,898	28.28%
ACEincome_2122 (Financial Hardship)			
Never	40,634	3,352	8.25%
Rarely	19,222	2,488	12.95%
Often	5,590	1,040	18.60%
Very Often	1,220	295	24.18%
ACESexDiscrim_2122 (Discrimination Based)			
No	65,597	6,854	10.45%
Yes	1,069	321	30.03%
ACEdivorce_2122 (Parental Separation)			
No	52,041	4,549	8.74%
Yes	14,625	2,626	17.95%
ACEdrug_2122 (Substance Abuse in Household)			
No	60,305	5,779	9.58%
Yes	6,361	1,396	21.95%
EmSFamily_2122 (Emotional Support from Family)			
Family	48,687	5,197	10.67%
Not Family	5,775	544	9.42%
EmSupport_2122 (General Emotional Support)			
Yes	54,462	5,741	10.55%
No	12,204	1,434	11.75%

Table 3.3 and Figure 3.6 signifies the overall and ADHD positive frequency and their distribution based on Adverse childhood experiences and social support factors. Considering the factors it can be stated that gender discrimination, parental separation, drugs, lack of family and emotional support have an impact on ADHD cases. ADHD rate jumps from 8.80% (none) to 28.28% (1+ ACEs), highlighting community adversity's impact. ADHD-positive rate is 30.03% for those experiencing discrimination, the highest among all factors.

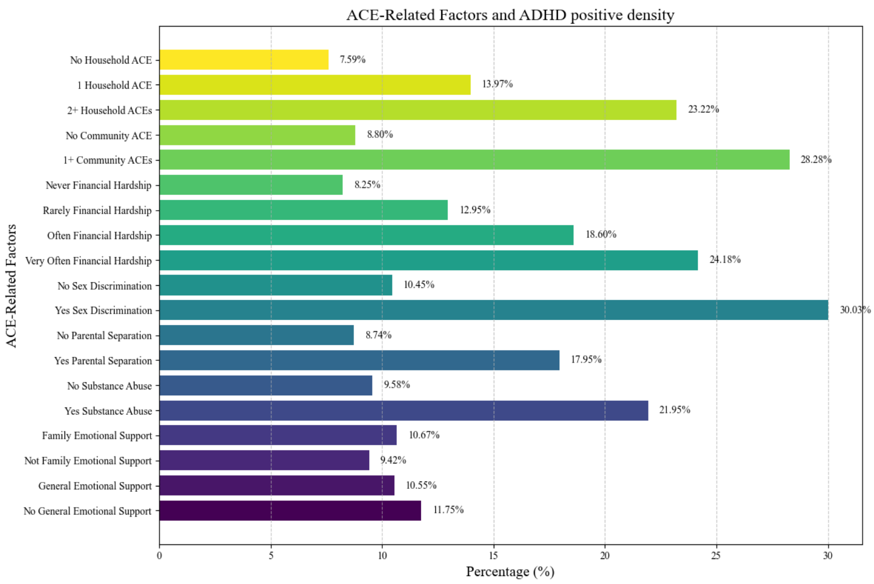

Figure 3.6: ACEs and ADHD positive density

Health and medical features

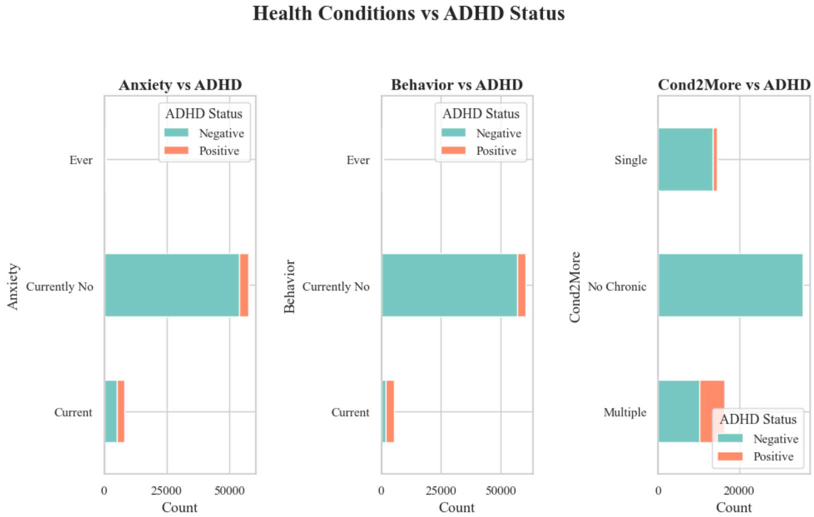

Figure 3.7: Top health contributors for ADHD

According to this study and survey, (anxiety, behavior, and chronic diseases) are the top health-contributing features for ADHD-positive cases. Figure 3.7 shows ADHD-free and ADHD-bearing ratios for those features. Having multiple health issues push towards ADHD. Severe behavior issues and ever had cases are also responsible for ADHD cases which is also the same for anxiety considering the Figure 3.7.

Environmental and lifestyle analysis

Table 3.4: Overall and ADHD Positive Frequency for Selected Lifestyle and Environment Features

Combinations	Overall Count	ADHD Positive	Percentage (%)
Fruit Intake (fruit_2122)			
2 times/day	54,658	6,916	12.65%
3+ times/day	3,330	67	2.01%
1 time/day	3,310	52	1.57%
4-6 times/month	3,266	72	2.20%
1-3 times/month	1,779	51	2.87%
Never	323	17	5.26%
Vegetable Intake (vegetables_2122)			
2 times/day	53,080	6,880	12.96%
3+ times/day	1,181	36	3.05%
1 time/day	3,747	62	1.65%
4-6 times/month	3,550	59	1.66%
1-3 times/month	4,069	85	2.09%
Never	1,039	53	5.10%
Sugary Drink Intake (SugarDrink_2122)			
No	56,249	6,960	12.38%
1-3 times/month	6,929	123	1.77%
4-6 times/month	1,441	35	2.43%
1 time/day	1,198	22	1.83%
2 times/day	618	20	3.24%
3+ times/day	231	15	6.49%
Breastfeeding History (BrstEver_2122)			
Yes	63,844	7,064	11.07%
No	2,822	111	3.93%
Exclusive Breastfeeding (ExBrstFd_2122)			
6+ months	54,404	6,900	12.69%
4-6 months	5,349	73	1.36%
4 months or less	4,091	91	2.22%
Never	2,822	111	3.93%
Park Access (park_2122)			
Yes	51,108	5,300	10.37%
No	15,558	1,875	12.05%
Library Access (library_2122)			
Yes	44,780	4,699	10.50%
No	21,886	2,476	11.32%
Smoking Household (smoking_2122)			
No	58,582	5,874	10.03%
Yes	8,084	1,301	16.09%

This part of features are interesting. Exploring these features it has been seen that most of the children are grown with great care. Table 3.4 shows the frequency of overall and ADHD-positive cases based on lifestyle and environmental features. The ratio of normal and ADHD cases are kind of similar for lifestyle and environmental features.

Behavioral and developmental exploration

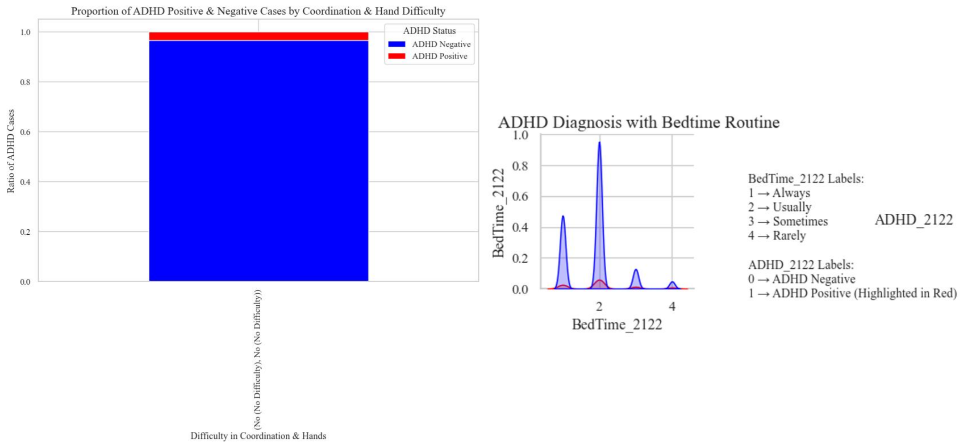

Figure 3.8: Bedroutine & Coordination impact on ADHD

Bed-routine, coordination and hand difficulty has impact on ADHD as shown in Figure 3.8. Most children follow a specific bed routine. Among them a portion has ADHD. But most child shows normal behavior. Among 23.90% children who don't follow bedtime get ADHD positive cases according to the survey. The coordination and hand difficulties have a mild impact on the ADHD.

3.1.3 Feature Selection Pipeline

Feature selection plays a critical role in improving model performance and reducing computational complexity, particularly when dealing with high-dimensional datasets. For this study, a multi-step feature selection strategy was employed to refine the initial dataset.

The process involved:

Correlation Matrix (Pearson) — to identify weakly correlated features with the target variable.

Variance Inflation Factor (VIF) — to detect and eliminate multicollinearity.

Recursive Feature Elimination (RFE) — to iteratively select the most relevant predictors based on model performance.

As illustrated in Figure 3.9, this pipeline ensured that only informative and non-redundant features were retained for model training. A total of 39 features with low predictive power or high multicollinearity were removed.

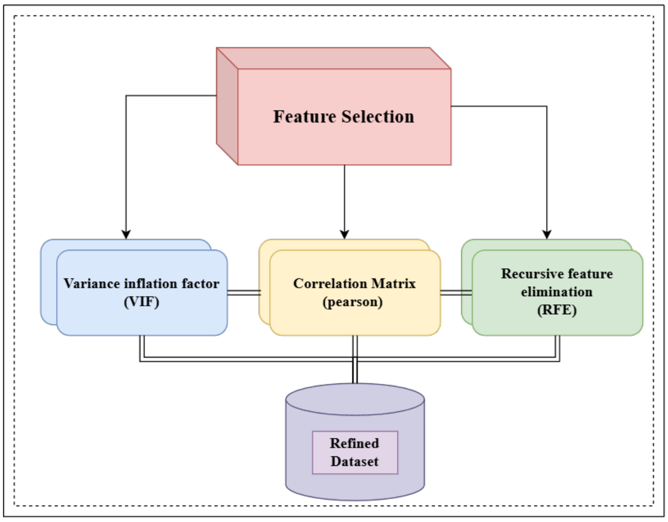

Figure 3.9: Feature selection workflow

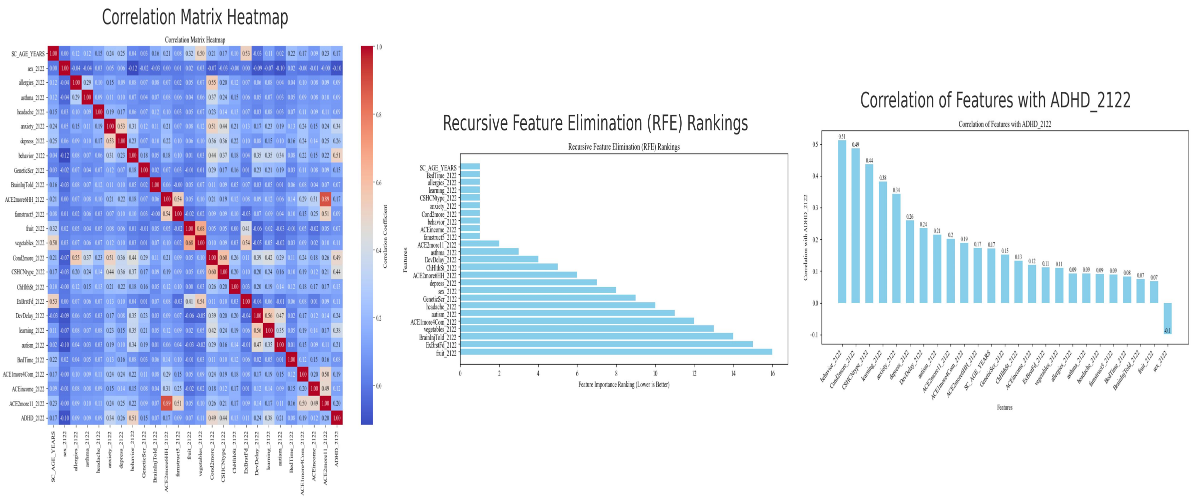

Figure 3.10: Selected Features Heatmap and Rankings

The refined dataset contains the mentioned features like the Table ?? and 26 features to be exact with 66,666 entries . The Figure 3.10 illustrates the correlation heatmap, recursive feature elimination rankings. Primarily, it can be stated that the behavior and chronic disease features put impact on the ADHD (Target) variable. For the RFE, the lower values are more impactful towards the target variable. VIF indicates the multicollinearity and the features that are more close to each other has been removed from the dataset for generating the precise and unbiased outcome.

Refined extracted dataset

Table 3.5: Final Refined Dataset

Feature Name	Relevant Question	Result Combination
SC_AGE_YEARS	What is the child's age?	Continuous
sex_2122	What is the child's gender?	Male, Female
allergies_2122	Has the child ever had allergies?	No, Ever, Current
asthma_2122	Has the child ever been diagnosed with asthma?	No, Ever, Current
headache_2122	Does the child experience headaches?	No, Ever, Current
anxiety_2122	Has the child ever been diagnosed with anxiety?	No, Ever, Current
depress_2122	Has the child ever been diagnosed with depression?	No, Ever, Current
behavior_2122	Does the child exhibit behavioral issues?	No, Ever, Current
GeneticScr_2122	Has the child undergone genetic screening?	Never, Ever, Identified
BrainInjTold_2122	Has the child been diagnosed with a brain injury?	Never, Ever, Had
ACE2more6HH_2122	How many adverse childhood experiences (household-based) has the child faced?	No, Single, Multiple
famstruct5_2122	What is the child's family structure?	2-parents, Single parent, Grandparent, Other
fruit_2122	How often does the child consume fruits?	No, 1-3x/m, 4-6x/m, 1x/day, 2x/day, 3+x/day
vegetables_2122	How often does the child consume vegetables?	No, 1-3x/m, 4-6x/m, 1x/day, 2x/day, 3+x/day
Cond2more_2122	How many health conditions does the child have?	None, Single, Multiple
CSHCNtype_2122	Does the child have special health care needs?	4 categories
ChHlthSt_2122	How is the child's overall health?	Good, Poor
ExBrstFd_2122	Was the child exclusively breastfed?	Never, Less than 6m, 6m, Greater than 6m
DevDelay_2122	Has the child been diagnosed with developmental delay?	No, Ever, Current
learning_2122	Does the child have a learning disability?	No, Ever, Current
autism_2122	Has the child been diagnosed with autism?	No, Ever, Current
BedTime_2122	Does the child have a consistent bedtime routine?	Always, Usually, Sometimes, Rarely
ACE1more4Com_2122	How many adverse childhood experiences (community-based) has the child faced?	No, Single, Multiple
ACEincome_2122	How often does the family struggle with income?	Never, Rarely, Somewhat, Very Often
ACE2more11_2122	Has the child experienced multiple adverse childhood experiences?	Single/None, Multiple
ADHD_2122 (Target)	Has the child been diagnosed with ADHD?	Negative, Positive

Total Frequency: 66,666

Table 3.5 picturize the portrait of the refined final dataset along with relevant questions, feature names and result combinations. The dataset size is 66,666 rows x 26 columns to be exact. It has been used for the ML,LLMs, XAI models training and evaluations. It holds a significant role in this study.

Data splitting

To prepare the data for model training and evaluation, the feature set was first separated from the target variable. This was accomplished by dropping the target column from the dataset to form the feature matrix, while the target variable was extracted from the corresponding column.

Subsequently, the dataset was split into training and testing subsets using the train_test_split function from the sklearn.model_selection module [34]. An 85:15 split ratio was applied, allocating 85% of the data for training and 15% for testing. To ensure reproducibility and consistent shuffling of the data, a fixed random seed of 42 was used.

This split strategy ensures an unbiased and balanced evaluation of the model's performance. The resulting training set comprised 56,666 rows $\times$ 25 features, while the test set included 10,000 rows $\times$ 25 features.

Feature scaling

To ensure that the features in the dataset are on a comparable scale and to improve the performance of the machine learning model, the Standard Scaler has been applied [35]. The Standard Scaler standardizes the features by transforming them such that they have a mean of 0 and a standard deviation of 1. This process ensures that each feature contributes equally to the model, preventing features with larger numerical ranges from disproportionately influencing the model's behavior. Standardization is particularly important for models that rely on distance metrics, such as linear regression, logistic regression, and support vector machines, as it ensures that no single feature dominates due to its scale. The given equation represents the standard scaler.

$$X_{\text{scaled}} = \frac{X - \mu}{\sigma} \tag{3.1}$$

3.1.4 Applied Machine Learning Algorithms

Fourteen machine learning models, including both basic and ensemble methods, were implemented to classify ADHD cases. Key mathematical formulations are included to support theoretical clarity.

Logistic Regression

A linear classification model using the sigmoid function to estimate class probabilities [36].

$$P(Y = 1|X) = \frac{1}{1 + e^{-z}}, \quad z = \beta_0 + \sum_{i=1}^{n} \beta_i X_i \tag{3.2}$$

Decision Tree

A rule-based learner that splits data using Gini impurity or Entropy to maximize information gain [37].

$$Gini = 1 - \sum_{i=1}^{n} p_i^2, \quad Entropy = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{3.3}$$

K-Nearest Neighbors (KNN)

A non-parametric algorithm that classifies based on majority vote among the k closest points [38].

$$y = \arg\max_{c \in C} \sum_{i=1}^{k} \nVdash(y_i = c) \tag{3.4}$$

$$d(x, x_i) = \sqrt{\sum_{j=1}^{n} (x_j - x_{i,j})^2} \tag{3.5}$$

Naïve Bayes

A probabilistic classifier based on Bayes' Theorem, assuming conditional independence between features [39].

$$P(C_k|X) = \frac{P(X|C_k)P(C_k)}{P(X)} \tag{3.6}$$

Linear SVC

A linear classifier that finds the hyperplane maximizing margin between classes [40]. Often used for binary classification and outlier detection.

Random Forest

An ensemble method that combines multiple decision trees trained on different data subsets to improve accuracy [41].

Gradient Boosting

A sequential technique where each learner tries to reduce the residual error of its predecessor [42].

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{3.7}$$

Extreme Gradient Boosting (XGBoost)

A scalable, regularized version of gradient boosting optimized for performance and speed [43, 44].

$$L_{xgb} = \sum_{i=1}^{N} L(y_i, F(x_i)) + \sum_{m=1}^{M} \Omega(h_m) \tag{3.8}$$

$$\Omega(h) = \gamma T + \frac{1}{2}\lambda |w|^2 \tag{3.9}$$

AdaBoost

Combines weak classifiers by assigning higher weights to misclassified instances [45].

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \tag{3.10}$$

LightGBM

An optimized gradient boosting method using histogram-based learning and leaf-wise tree growth [46].

$$F_m(x) = F_{m-1}(x) + \sum_{j=1}^{J_m} w_j h_j(x) \tag{3.11}$$

CatBoost

A gradient boosting algorithm developed for categorical feature handling, using ordered boosting to prevent overfitting [47].

$$G_t(x) = G_{t-1}(x) + \alpha \sum_{k=1}^{K_t} v_k g_k(x) \tag{3.12}$$

Stacking

An ensemble method that combines predictions of multiple base models using a meta-learner [48]. Improves accuracy by learning how to best weight each model's output.

Blending (ML)

Similar to stacking, blending uses a holdout validation set to train a meta-model, often using logistic regression [49].

Blending (DL)

An ensemble of deep learning architectures (e.g., CNN, RNN, Transformer) combined via a meta-model. Helps generalize predictions while minimizing overfitting [50].

Hyperparameter Optimization

Hyperparameter optimization is required to improve the performance of machine learning models, especially in applications such as ADHD prediction, where reliability and accuracy are necessary. In this study, we used both GridSearchCV and Optuna for optimizing the hyperparameters of various machine learning algorithms. These methods systematically look for the best hyperparameter configurations in order to enhance model performance. Figure 3.11 illustrates the used hyperparameter tuning techniques for machine learning models.

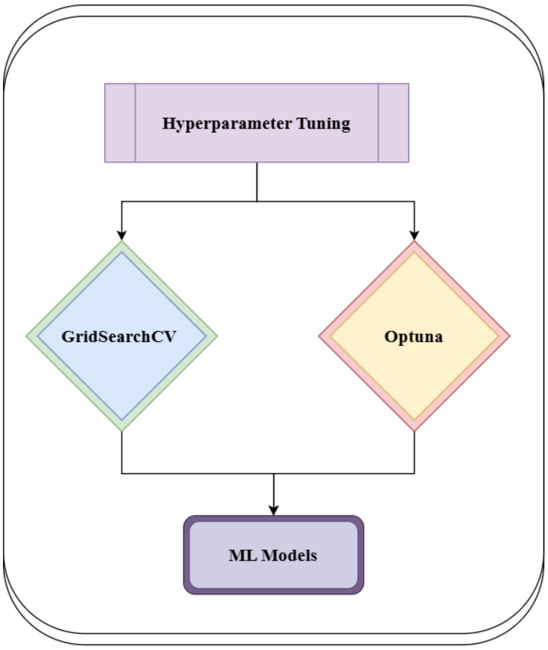

Figure 3.11: Hyperparameter tuners for the ML models

3.1.5 Applied Transformer Models(LLMs)

To enhance ADHD prediction through natural language understanding, several BERT-based transformer models were employed. These models process medical and survey text data with high accuracy and contextual awareness.

BERT and Variants

BERT (Bidirectional Encoder Representations from Transformers) [51] serves as the foundation for language understanding, pre-trained on 3.3B words from Wikipedia and BooksCorpus. Its lighter variant, DistilBERT [52], retains 97

ClinicalBERT [53] and BioBERT [54] are domain-specific adaptations fine-tuned for clinical and biomedical texts, improving tasks like NER, relation extraction, and medical classification. ALBERT [55] further reduces parameter size via matrix factorization and cross-layer sharing, improving efficiency for large-scale NLP tasks.

LLMs Workflow

As illustrated in Figure 3.12, the processed dataset was further refined for LLMs by performing tokenization, text encoding, and data splitting. This setup enables BERT-based models to extract meaningful patterns for ADHD-related predictions.

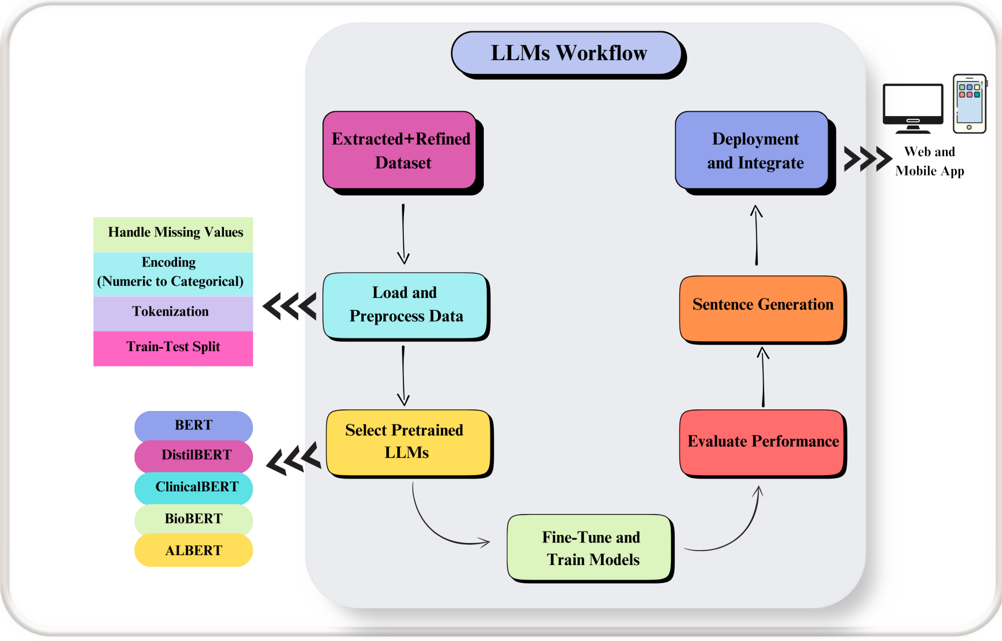

Figure 3.12: LLMs workflow

All models were evaluated for accuracy, inference time, and generalization. DistilBERT outperformed other variants with the highest accuracy and fastest runtime (see Table 4.7), demonstrating its practical advantage for deployment.

3.1.6 Dialogflow-based therapist Chatbot "Comfort Chat"

The proposed system includes a lightweight and user-friendly chatbot named ComfortChat, developed using Google's Dialogflow platform [56]. Designed to offer primary therapeutic support, ComfortChat enhances user engagement by delivering quick, relevant responses.

Dialogflow is widely adopted in healthcare for building virtual assistants, symptom checkers, and patient support tools. It leverages natural language understanding (NLU) to identify medical intents and extract key entities such as symptoms and conditions. Its integration capabilities, multilingual support, and compliance with healthcare standards make it ideal for real-time medical interaction.

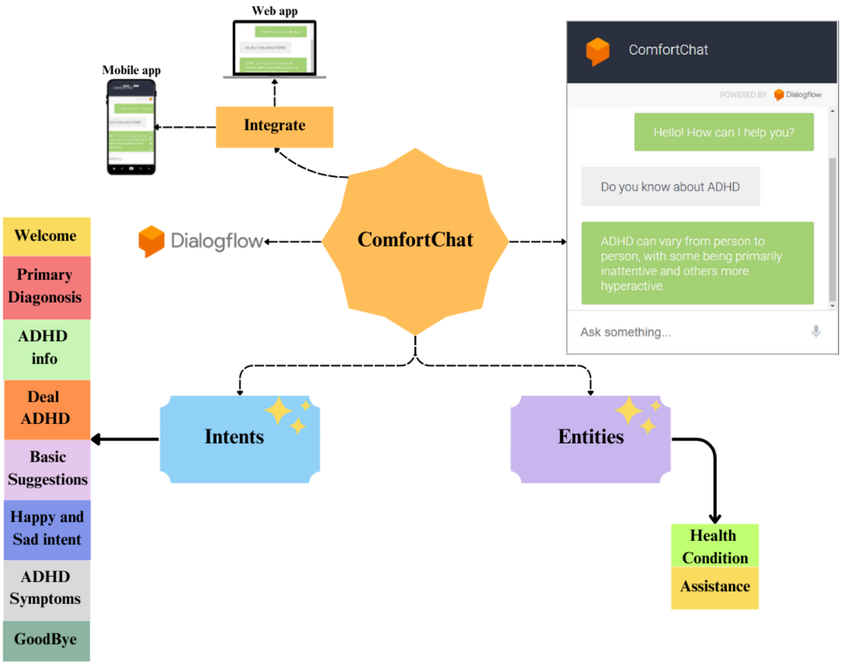

Figure 3.13: Structure of Chatbot (ComfortChat)

Structure of ComfortChat

ComfortChat is built using Dialogflow and integrated with both web and mobile apps for user accessibility. The system is driven by two core components: Intents and Entities.

- **Intents:** Guide responses based on user queries. Examples include: Welcome, Diagnosis, ADHD Info, Suggestions, Emotions, and Goodbye.
- **Entities:** Extract key data like *Health Condition* and *Assistance* to make replies more relevant.

This structure enables ComfortChat to deliver helpful, context-aware support for ADHD-related concerns across platforms.

The overall pipeline spans from data collection to deployment, incorporating rigorous processing at every stage.

Chapter 4

Experiment Result Analysis and Discussion

Baseline performance (without resampling)

The baseline without applying any resampling techniques provide unbiased and impartial outcomes. Fourteen ML model algorithms have been trained without applying any resampling methods.

Table 4.1: Performance Metrics of Machine Learning and Other Models

Type	**Model**	**Acc (%)**	**Pre (%)**	**Rec (%)**	**F1 (%)**
Basic ML Models	Logistic Regression	92.54 ± 0.21	82.23 ± 0.69	75.34 ± 0.78	78.22 ± 0.70
	Decision Trees	91.34 ± 0.29	77.41 ± 0.75	77.31 ± 0.91	77.36 ± 0.80
	Linear SVC	92.50 ± 0.23	82.63 ± 0.88	74.21 ± 0.62	77.58 ± 0.67
	k-NN	92.32 ± 0.12	81.31 ± 0.54	75.31 ± 0.47	77.87 ± 0.31
	Naïve Bayes	85.72 ± 0.39	69.31 ± 0.41	82.00 ± 0.37	72.98 ± 0.49
Ensemble Models	Random Forest	93.29 ± 0.15	83.36 ± 0.50	79.99 ± 0.80	81.54 ± 0.50
	AdaBoost	92.46 ± 0.18	81.80 ± 0.62	75.60 ± 0.78	78.24 ± 0.61
	XGBoost	93.80 ± 0.13	85.01 ± 0.70	81.05 ± 0.41	82.86 ± 0.14
	LightGBM	93.79 ± 0.09	86.00 ± 0.61	79.24 ± 0.55	82.14 ± 0.23
	Gradient Boosting	93.23 ± 0.07	84.35 ± 0.42	77.49 ± 0.54	80.40 ± 0.27
	CatBoost	93.59 ± 0.06	85.50 ± 0.51	78.49 ± 0.55	81.48 ± 0.24
Blending & Stacking	Blending (ML Models)	93.44 ± 0.03	83.84 ± 0.11	81.40 ± 0.11	82.56 ± 0.06
	Stacking	**94.27 ± 0.02**	**87.00 ± 0.15**	**82.07 ± 0.08**	**84.30 ± 0.03**
	Blending (DL Models)	93.49 ± 0.03	83.94 ± 0.06	81.64 ± 0.14	82.73 ± 0.10

Among the fourteen ML models the stacking ensemble comes out with the best performance having 94.27% accuracy, 87.00% precision, 82.07% recall and 84.30% f1 score. Table 4.1 portrays the baseline performances of the fourteen ML algorithms of the proposed framework.

Performance with oversampling(SMOTE)

Synthetic Minority Oversampling Technique is referred to as SMOTE. Instead of replicating pre-existing cases, this widely used oversampling technique generates synthetic examples. For analyzing the oversampling outcomes all the algorithms have been trained applying SMOTE.

Table 4.2: Performance Metrics of Machine Learning Models with SMOTE Resampling

Type	**Model**	**Acc (%)**	**Pre (%)**	**Rec (%)**	**F1 (%)**
Basic ML Models	Logistic Regression	86.32	71.00	88.00	75.00
	Decision Trees	90.97	76.00	78.00	77.00
	Linear SVC	85.93	71.00	88.00	75.00
	k-NN	88.98	73.00	85.00	77.00
	Naïve Bayes	84.13	68.00	84.00	72.00
Ensemble Models	Random Forest	92.78	**81.00**	81.00	82.00
	AdaBoost	89.35	74.00	85.00	78.00
	XGBoost	**92.92**	**81.00**	84.00	**83.00**
	LightGBM	91.88	78.00	87.00	82.00
	CatBoost	91.39	77.00	**89.00**	82.00
	Gradient Boosting	92.76	80.00	83.00	82.00
Blending & Stacking	Stacking Model	92.86	**81.00**	82.00	81.00
	Blending Model	92.82	**81.00**	83.00	82.00

From Table 4.2 it can be stated that the highest accuracy of 92.92% is achieved by xgboost classifier. The top precision score is 81.00% for four different models (xgboost, random forest, stacking and blending). Conversely, the CatBoost classifier achieved the highest recall score of 89.00%. The xgboost also holds the top f1 score of 83.00%.

Performance with undersampling

The undersampling technique makes the unbalanced dataset balanced by reducing the number of samples from the majority class. This technique has been applied to observe the results by different ML algorithms.

Table 4.3: Performance Metrics of Machine Learning and Other Models with Undersampling

Type	Model	Acc (%)	Pre (%)	Rec (%)	F1 (%)
Basic ML Models	Logistic Regression	85.70	71.00	88.00	75.00
	Decision Trees	88.39	73.00	87.00	77.00
	Linear SVC	85.19	70.00	88.00	74.00
	k-NN	86.20	71.00	87.00	75.00
	Naïve Bayes	84.42	69.00	84.00	72.00
Ensemble Models	Random Forest	87.91	73.00	90.00	78.00
	AdaBoost	86.23	71.00	89.00	75.00
	XGBoost	88.19	73.00	91.00	78.00
	LightGBM	88.24	74.00	91.00	78.00
	CatBoost	87.40	73.00	**92.00**	77.00
	Gradient Boosting	88.19	73.00	91.00	78.00
Blending & Stacking	Stacking	88.88	74.00	91.00	79.00
	Blending	**89.79**	**75.00**	89.00	**80.00**

Table 4.3 provides the performances of all ML algorithms for the proposed framework. Interestingly, it has been observed that the blending ensemble performs the best with 89.79% accuracy, 75.00% precision, and 80.00% f1 score. However, the highest recall score has been achieved by the catboost classifier of 92.00%.

Hyperparameter Optimization and Analysis

Table 4.4: Hyperparameter Tuning Results: With Best Parameters

Method	Model	Accuracy (%)	Best Parameters
GridSearch	Logistic Regression	92.63	{model__C: 0.01, model__solver: 'lbfgs'}
	Decision Trees	92.90	{model__max_depth: 10, model__min_samples_split: 5}
	Linear SVC	92.50	{model__C: 1.0, model__kernel: 'linear'}
	k-NN	92.40	{model__n_neighbors: 9}
	Naïve Bayes	85.72	{alpha: 0.1, fit_prior: True}
	Random Forest	93.64	{model__max_depth: 20, model__n_estimators: 500}
	AdaBoost	92.56	{model__learning_rate: 0.1, model__n_estimators: 100}
	XGBoost	93.80	{model__learning_rate: 0.2, model__n_estimators: 100}
	LightGBM	93.86	{model__learning_rate: 0.1, model__n_estimators: 100}
	Gradient Boosting	93.81	{model__learning_rate: 0.2, model__n_estimators: 300}
	CatBoost	93.87	{model__depth: 6, model__iterations: 300, model__learning_rate: 0.1}
	Stacking (ML models)	94.20	{rf_n_estimators: 150, xgb_learning_rate: 0.05, meta_C: 0.03}
	Blending (ML models)	93.90	{log_C: 0.01, knn_n_neighbors: 9, rf_n_estimators: 300}
Optuna	Logistic Regression	92.75	{C: 0.0027, solver: 'lbfgs'}
	Decision Trees	93.23	{max_depth: 9, min_samples_split: 7}
	Linear SVC	92.70	{C: 0.03, kernel: 'linear'}
	k-NN	92.48	{n_neighbors: 15}
	Naïve Bayes	85.92	{alpha: 0.13, fit_prior: True}
	Random Forest	93.92	{n_estimators: 432, max_depth: 20}
	AdaBoost	92.68	{n_estimators: 131, learning_rate: 0.0589}
	XGBoost	93.95	{learning_rate: 0.0912, n_estimators: 261}
	LightGBM	94.06	{learning_rate: 0.0516, n_estimators: 315}
	Gradient Boosting	94.00	{learning_rate: 0.1093, n_estimators: 420}
	CatBoost	94.14	{learning_rate: 0.0374, depth: 7, iterations: 391}
	Stacking (ML models)	94.30	{rf_n_estimators: 200, lgbm_n_estimators: 250, xgb_n_estimators: 500, xgb_learning_rate: 0.0482, xgb_max_depth: 5, meta_C: 0.0357}
	Blending (ML models)	94.00	{log_C: 0.02, knn_n_neighbors: 11, rf_n_estimators: 350}

Here, Table 4.4 outlines the best outcome of the ML models with the best parameters. Considering the accuracy, clearly the stacking baseline ensemble outshines all other ML algorithms with quite a margin. Both optuna and gridsearch confirm the statement. Optuna outperforms gridsearch by using intelligent, adaptive sampling strategies that efficiently explore the hyperparameter space, significantly reducing computation time while achieving better model performance. For this reason, the parameters the optuna provided were chosen and utilized for the suggested system. The params are {'xgb_n_estimators': 500, 'xgb_learning rate': 0.0482, 'xgb_max depth': 5, 'meta_C': 0.0357} .

Confusion matrix

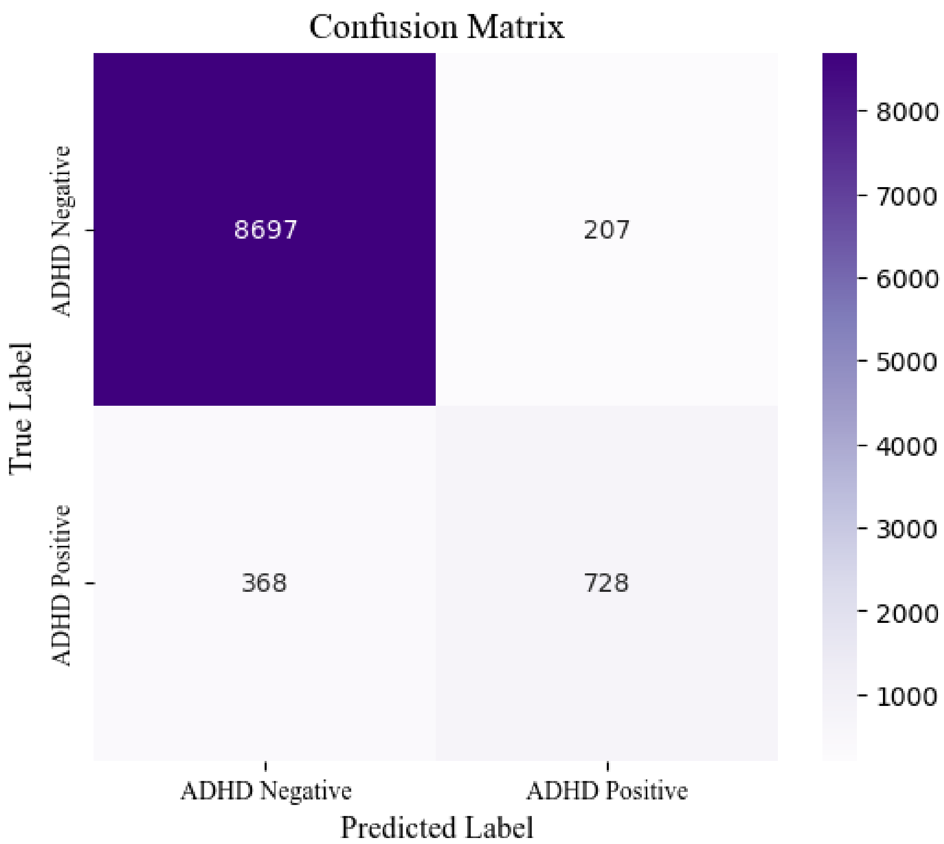

Figure 4.1: Confusion Matrix for the Proposed Model(Stacking Baseline)

Figure 4.1 states the confusion matrix for the proposed stacking model. The model holds a strong macro recall of 82.07%, demonstrating the model's generalizability to both classes. The recall for the ADHD-positive cases demonstrates that the model can identify a very high percentage of actual ADHD cases without even employing resampling techniques. Also, the strong recall of the ADHD-negative cases demonstrates that the model is effective at identifying individuals without

ADHD, and there are few false positives. As the data is in its raw form, it provides unbiased outcomes.

ROC AUC curves

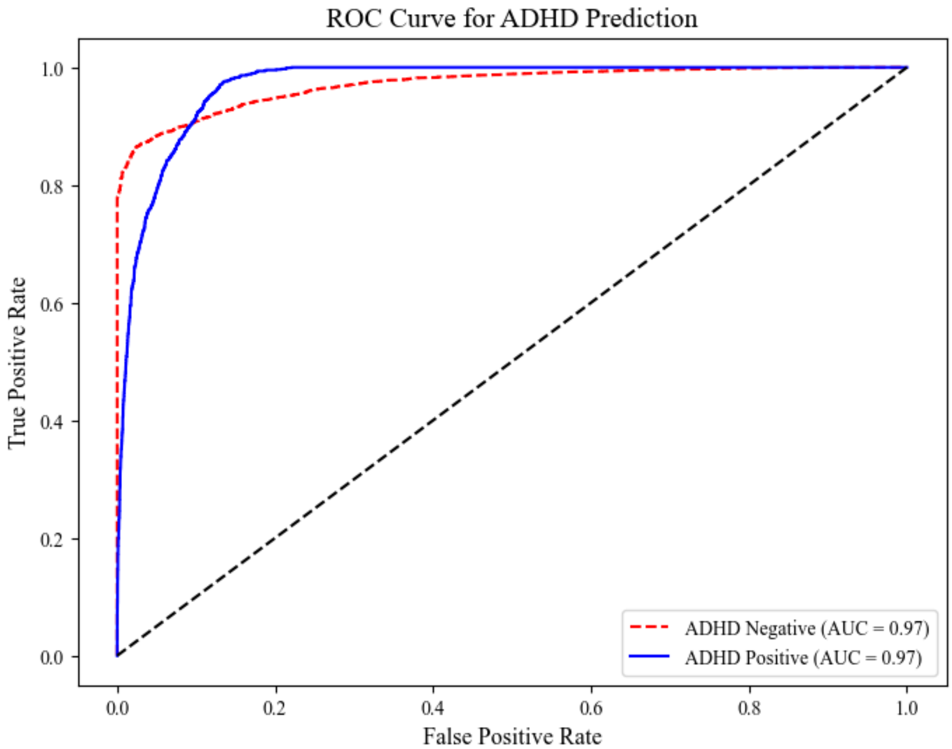

Figure 4.2: ROC-AUC curve for the Proposed Model(Stacking Baseline)

The ROC-AUC curve from Figure 4.2 evaluates the performance of the stacking baseline model in distinguishing between ADHD-positive and ADHD-negative cases. Both classes achieve a score of 0.97 which indicates excellent discriminatory ability of the model. The model's durability is demonstrated by its strong performance in both classes, even in the absence of resampling. The ROC curve for both classes is sharply increasing at the beginning, which means that the model has high true positive rates with low false positives, which is desirable for classification problems. AUC values close to 1.0 mean that the model is good at distinguishing between positive and negative cases.

Precision Recall curves

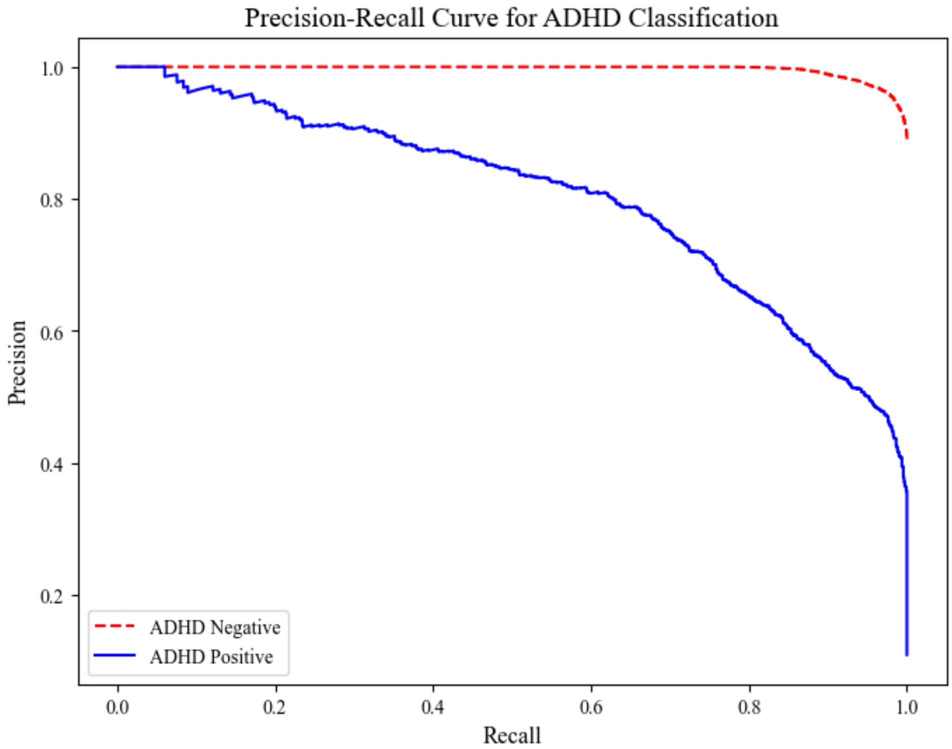

Figure 4.3: Precision-Recall curve for the Proposed Model(Stacking Baseline)

The Precision-Recall (PR) curve provides a detailed view of how well the stacking ensemble balances precision and recall for ADHD classification. Figure 4.3 outlines the precision-recall curve for the proposed framework. It has been observed that the model gains 87.00% precision and 82.07% recall values. Both positive and negative class outcomes are well-balanced and maintain a strong overall performance without resampling techniques.

Learning curves

The learning curve provides valuable insights into the stacking baseline model's performance as the training size increases, helping to assess bias-variance tradeoffs and model generalization.

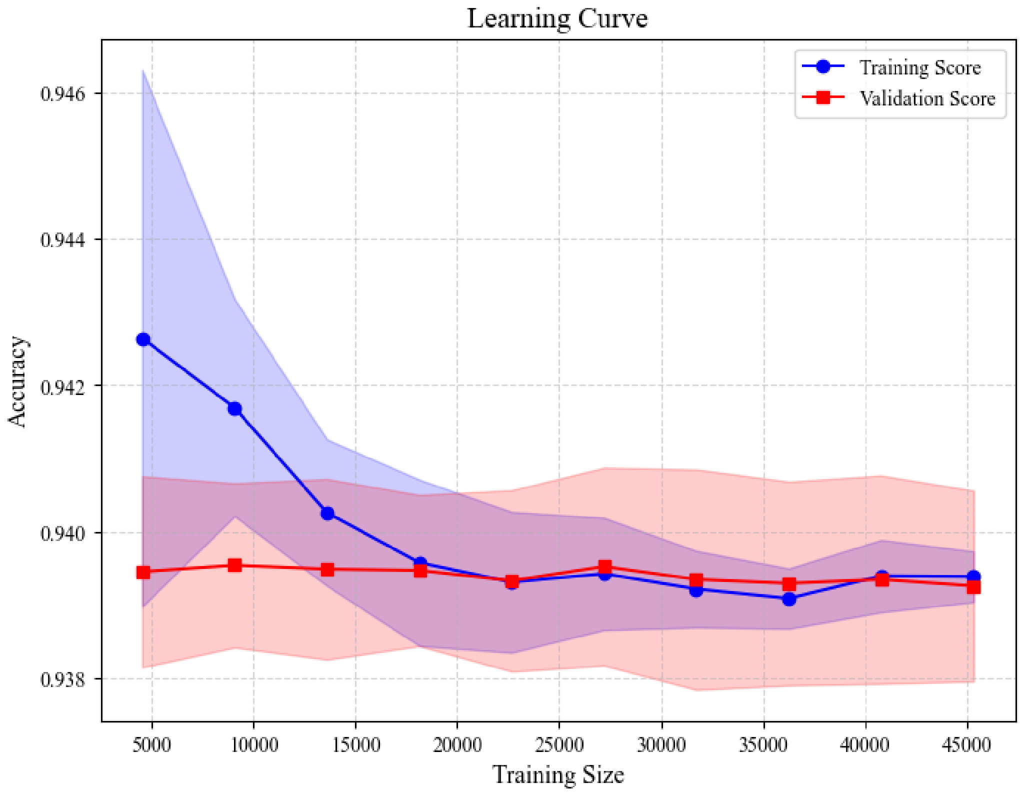

Figure 4.4: Learning curve for the Proposed Model(Stacking Baseline)

Initially, the model shows high variance as it is working on the smaller datasets but as more and more data flow into the system, the variance reduces while the validation accuracy levels off 94.00%. Not having a huge difference between validation and training scores indicates that neither the model is overfitting nor underfitting, and it is best optimized for the classification of ADHD and the Figure 4.4 outlines that.

Ablation studies

Ablation experiments [57] remove or modify systematically some features or elements of a model to quantify their contribution separately to performance. Ablation is employed to identify the most impactful features and enhance model complexity in the optimal manner.

Table 4.5 shows the ablation study results for the proposed study. Removing 5 less important feature based on feature importance provides the best outcome. It can be observed that the performance gradually declines as the number of features decreases.

Table 4.5: Performance metrics of Machine Learning and other models with feature ablation

Num. of Features	Removed Features	Accuracy	Precision	Recall	F1 Score
20	5	**0.9423**	**0.8674**	**0.8144**	**0.8381**
25	0	0.9417	0.8652	0.8135	0.8367
24	1	0.9413	0.8647	0.8114	0.8352
23	2	0.9419	0.8665	0.8130	0.8369
22	3	0.9416	0.8656	0.8122	0.8361
21	4	0.9423	0.8675	0.8141	0.8380
19	6	0.9407	0.8638	0.8083	0.8330
18	7	0.9390	0.8598	0.8022	0.8277
17	8	0.9390	0.8606	0.8010	0.8272
16	9	0.9370	0.8545	0.7953	0.8213
15	10	0.9365	0.8517	0.7959	0.8206
14	11	0.9361	0.8498	0.7963	0.8201
13	12	0.9360	0.8507	0.7941	0.8190
12	13	0.9344	0.8452	0.7904	0.8146
11	14	0.9322	0.8431	0.7760	0.8047
10	15	0.9294	0.8385	0.7603	0.7926

4.0.1 Performance evaluation metrics(LLMs)

Transformer Model Evaluation Measures(BERT)

BERT transformer-based models, designed for NLP and highly prevailing NLU, form the foundation architecture for many powerful Large Language Models (LLMs) by enabling deep contextual human language comprehension. In this particular study, five different BERT-based transformed models have been trained for comprehensive experiments. It consists of BERT, DistilBERT, ClinicalBERT, BioBERT, and ALBERT models. These models were selected to represent a diverse range of general-purpose, distilled, clinical, biomedical, and lightweight transformer architectures. Each model was fine-tuned on the collected ADHD-related dataset to evaluate their performance in early prediction tasks. Special emphasis was placed on comparing their interpretability, efficiency, and classification accuracy. The findings provide valuable insights into the applicability of domain-specific BERT variants in mental health-related predictions. Overall, the BERT family models served as a crucial benchmark to complement the machine learning ensemble models explored in the study.

Table 4.6: Evaluation Metrics of Transformer Models (BERT)

Model	Validation Loss	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	AUC	Runtime (s)
BERT	0.247	87.15	91.52	87.15	88.62	0.911	36.47
Epoch 1	0.247	87.14	91.52	87.14	88.62	0.910	-
Epoch 2	0.235	89.19	79.55	89.19	84.10	0.817	-
Epoch 3	0.237	89.19	79.55	89.19	84.10	0.919	-
DistilBERT	0.142	93.45	93.22	93.45	93.32	0.968	19.25
Epoch 1	0.158	92.78	93.20	92.78	92.96	0.960	-
Epoch 2	0.162	93.34	92.78	93.34	92.77	0.965	-
Epoch 3	0.142	93.45	93.22	93.45	93.32	0.968	-
ClinicalBERT	0.343	89.27	90.42	89.27	84.28	0.291	36.04
Epoch 1	0.343	89.27	90.42	89.27	84.28	0.291	-
Epoch 2	0.343	89.19	79.55	89.19	84.10	0.893	-
Epoch 3	0.341	89.19	79.55	89.19	84.10	0.913	-
BioBERT	0.344	89.19	79.55	89.19	84.10	0.731	42.17
Epoch 1	0.344	89.19	79.55	89.19	84.10	0.731	-
Epoch 2	0.252	89.19	79.55	89.19	84.10	0.902	-
Epoch 3	0.236	89.19	79.55	89.19	84.10	0.909	-
ALBERT	0.183	92.70	92.21	92.70	92.38	0.950	46.25
Epoch 1	0.240	87.14	91.52	87.14	88.61	0.905	-
Epoch 2	0.224	92.39	91.75	92.39	91.94	0.911	-
Epoch 3	0.182	92.70	92.21	92.70	92.38	0.950	-

Table 4.6 outlines the evaluation metrics of the BERT-based transformer models. Each transformer model has been evaluated over the course of three training epochs, allowing for a comprehensive assessment of performance progression across training iterations.

Table 4.7: Final Comparison of Transformer Models

Model	Loss	Acc (%)	Pre (%)	Rec (%)	F1 (%)	AUC	Runtime (s)
BERT	0.247	87.15	91.52	87.15	88.62	0.911	36.47
DistilBERT	**0.142**	**93.45**	**93.22**	**93.45**	**93.32**	**0.968**	**19.25**
ClinicalBERT	0.343	89.27	90.42	89.27	84.28	0.291	36.04
BioBERT	0.344	89.19	79.55	89.19	84.10	0.731	42.17
ALBERT	0.183	92.70	92.21	92.70	92.38	0.950	46.25

Summarizing the Table 4.6 , the Table 4.7 has been retrieved that outlines the final hypertuned optimized results. It also confirms that DistilBERT provides the best outcome with a loss of 0.142, runtime of 19.25s, and accuracy of 93.45%. Also has a high precision, recall and f1 score.

Comparison graphs for BERT models

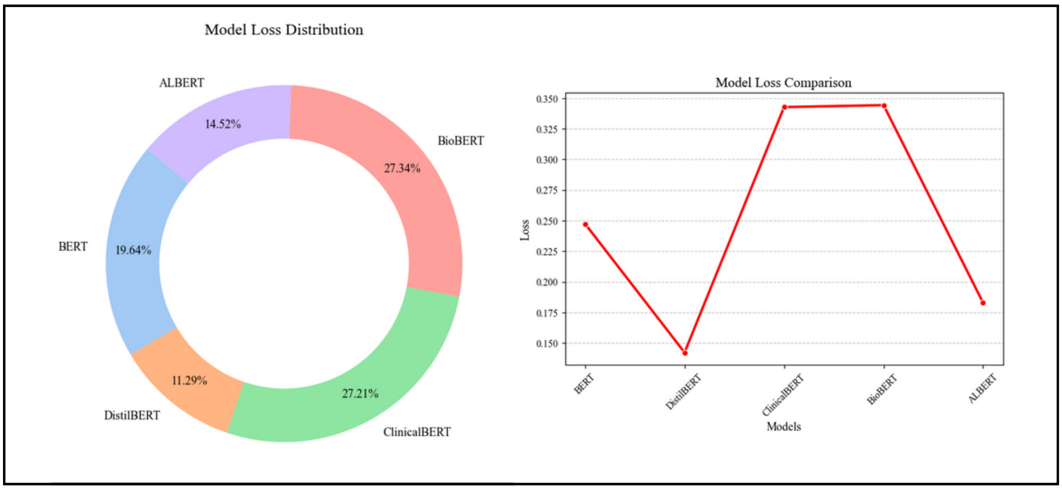

Figure 4.5: Model Loss Distribution(LLMs)

Analyzing Figure 4.5 , it can be stated that DistilBERT has the lowest validation loss, according to the model loss distribution, indicating better generalization and optimization. ClinicalBERT and BioBERT, on the other hand, show the greatest losses, indicating comparatively poorer performance across the models that were assessed.

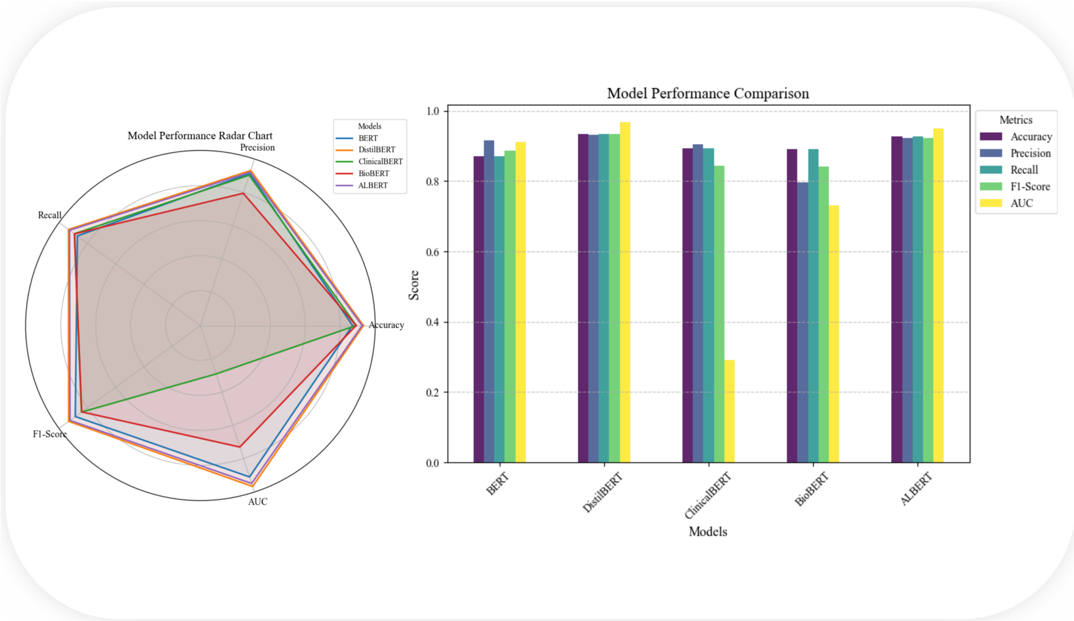

Figure 4.6: Model Comparison with Radarchart (LLMs)

DistilBERT continues to outperform other models in all the metrics like accuracy, precision, recall, F1-score, and AUC, as can be seen from the radar and bar plots of

Figure4.6. ClinicalBERT's overall performance curve is impacted by its enormous AUC gap. In important metrics, BERT and ALBERT's performance is on par with DistilBERT, but marginally worse.

4.0.2 Generalization and robustness

Performance on 2018–2019 National Survey of Children's Health Dataset

The proposed system was evaluated using the 2018–2019 NSCH dataset [58], with five features removed for alignment. Table 4.8 presents the results, where the stacking ensemble model achieved the highest accuracy and precision.

Table 4.8: Performance on 2018–2019 NSCH Dataset (Test Set 1)

Type	**Model**	**Acc (%)**	**Pre (%)**	**Rec (%)**	**F1 (%)**
Basic ML Models	Logistic Regression	92.57 ± 0.18	82.33 ± 0.55	75.61 ± 0.57	78.44 ± 0.57
	Decision Trees	91.57 ± 0.11	78.42 ± 0.35	78.01 ± 0.44	78.16 ± 0.48
	Linear SVC	92.51 ± 0.17	82.62 ± 0.62	74.48 ± 0.46	77.77 ± 0.50
	k-NN	92.58 ± 0.18	81.85 ± 0.57	76.86 ± 0.81	79.06 ± 0.63
	Naïve Bayes	86.24 ± 0.38	69.90 ± 0.49	**83.31 ± 0.67**	73.66 ± 0.59
Ensemble Models	Random Forest	93.36 ± 0.18	83.51 ± 0.57	80.47 ± 0.59	81.89 ± 0.50
	AdaBoost	92.54 ± 0.16	81.99 ± 0.50	76.09 ± 0.55	78.63 ± 0.49
	XGBoost	93.73 ± 0.13	84.60 ± 0.47	81.44 ± 0.43	82.92 ± 0.34
	LightGBM	93.91 ± 0.13	85.21 ± 0.51	81.69 ± 0.37	**83.32 ± 0.31**
	Gradient Boosting	93.68 ± 0.12	85.01 ± 0.53	80.21 ± 0.73	82.36 ± 0.43
	CatBoost	93.67 ± 0.18	85.47 ± 0.64	79.26 ± 0.70	81.96 ± 0.55
Blending & Stacking	Blending (ML models)	93.44 ± 0.08	84.29 ± 0.16	76.13 ± 0.32	79.49 ± 0.20
	Stacking (ML models)	**94.27 ± 0.08**	**86.08 ± 0.40**	80.16 ± 0.72	82.78 ± 0.38

The stacking model achieved the best accuracy (94.27%) and precision (86.08%), showcasing excellent performance and robustness during generalization on past-year survey data.

Performance on 2022–2023 National Survey of Children's Health Dataset

Further evaluation was conducted using the latest 2022–2023 NSCH dataset [58]. As shown in Table 4.9, the blending ensemble model outperformed all others.

Table 4.9: Performance on 2022–2023 NSCH Dataset (Test Set 2)

Type	Model	Acc (%)	Pre (%)	Rec (%)	F1 (%)
Basic ML Models	Logistic Regression	92.62 ± 0.16	82.50 ± 0.51	75.73 ± 0.44	78.58 ± 0.47
	Decision Trees	91.43 ± 0.21	77.85 ± 0.45	77.04 ± 0.48	77.48 ± 0.44
	Linear SVC	92.51 ± 0.16	82.68 ± 0.63	74.40 ± 0.39	77.73 ± 0.45
	k-NN	92.44 ± 0.19	81.39 ± 0.52	76.59 ± 0.81	78.71 ± 0.67
	Naïve Bayes	86.45 ± 0.39	70.08 ± 0.52	83.22 ± 0.66	73.84 ± 0.61
Ensemble Models	Random Forest	93.04 ± 0.17	82.62 ± 0.61	79.64 ± 0.39	81.03 ± 0.39
	AdaBoost	92.55 ± 0.14	82.06 ± 0.45	76.07 ± 0.56	78.64 ± 0.45
	XGBoost	93.53 ± 0.11	84.17 ± 0.43	80.56 ± 0.57	82.23 ± 0.34
	LightGBM	93.78 ± 0.08	84.98 ± 0.33	81.09 ± 0.54	82.87 ± 0.29
	Gradient Boosting	93.64 ± 0.08	84.81 ± 0.24	80.20 ± 0.84	82.27 ± 0.45
	CatBoost	93.60 ± 0.14	85.18 ± 0.62	79.22 ± 0.65	81.82 ± 0.44
Blending & Stacking	**Blending (ML models)**	**94.98 ± 0.07**	**89.14 ± 0.12**	**86.25 ± 0.33**	**87.62 ± 0.19**
	Stacking (ML models)	93.88 ± 0.18	88.20 ± 0.53	80.73 ± 0.89	83.91 ± 0.63

The blending model surpassed all others, reaching an accuracy of 94.98%, precision of 89.14%, and F1-score of 87.62%. This marks a significant leap in performance over the stacking model, reflecting the proposed system's ability to generalize across independent, unseen datasets.

4.0.3 (Clinical/Domain)Expert assessment and feedback

The insights and observations made by domain professionals, validate the practical appplicability and usefulness of the proposed system in real-world scenarios and deployment. Also, feedback reveals significant strengths, potential limitations, and actionable recommendations for further refinement. To ensure clarity and credibility, we sought feedback from institutions and professionals, including the National Institute of Mental Health (NIMH), Bangladesh, and a child specialist from Enam Medical College. Insights from psychiatrists, caregivers, parents, and AI researchers helped validate the system and guide improvements. This collaborative input aligns the solution with real-world clinical needs, and future development will continue to incorporate such feedback for ongoing refinement.

4.1 Model Interpretability

Overview of XAI

Explainable AI stands out with the interpretations and insights. They are often responsible for extracting the feature importance. LIME and SHAP has been used for the proposed modelling.

LIME outcomes

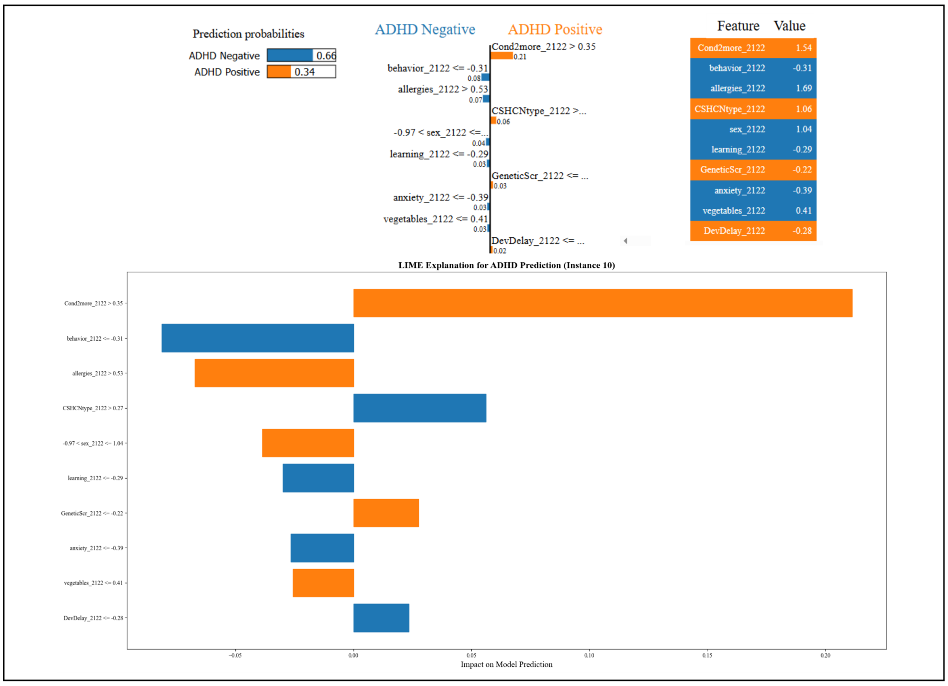

Figure 4.7: LIME explanation

LIME is a method that uses a local interpretable model to explain every single prediction, which can model any black box machine learning model. The model is estimating a 66% probability of the instance being ADHD Negative and a 34% probability of being ADHD Positive as shown in the Figure 4.7. The model predicts "ADHD Positive" if the person has more than one health condition (Cond2more_2122), allergies, or a particular type of CSHCN, showing that these are significant risk factors. Behavioral (behavior_2122), anxiety, eating (vegetables_2122), and development signs (DevDelay_2122, learning_2122) appear to reduce the chances of predicting ADHD in this instance.

Table 4.10: LIME Explanation for ADHD Prediction (Instance 10)

Feature	Value	Impact on ADHD Prediction
Cond2more_2122	1.54	Strongest positive influence, significantly increasing probability
allergies_2122	1.69	Moderate positive influence
CSHCNtype_2122	1.06	Mild positive influence
GeneticScr_2122	-0.22	Small positive and slight negative influences
DevDelay_2122	-0.28	Minor positive and slight negative influences
behavior_2122	-0.31	Strongest negative influence, significantly decreasing probability
anxiety_2122	-0.39	Moderate negative influence
vegetables_2122	0.41	Moderate negative influence
learning_2122	-0.29	Slight negative influence

The Table 4.10 illustrates the LIME Explanation for ADHD Prediction. It also picturize that children having multiple chronic diseases and severe behavioral issues have most impact on ADHD positive and reverse cases for the ADHD negetive.

SHAP outcomes

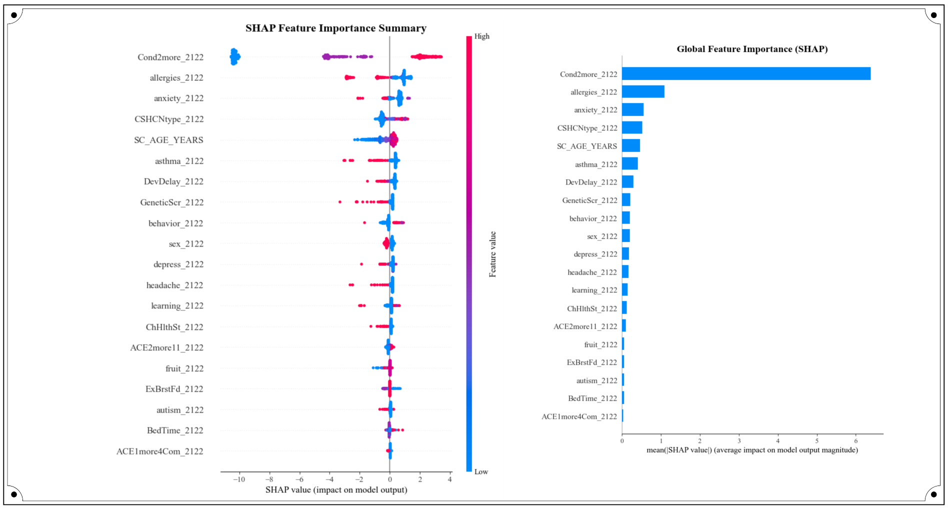

Figure 4.8: SHAP explanation

Through Figure 4.8, SHAP also confirms the (Cond2more_2122) is the most influential feature for the ADHD positive cases in children. Its strong and steady correlation with ADHD outcomes is indicated by its high SHAP scores. (allergies_2122,anxiety_2122) also hold substantial global influence. The limited global influence of features such as sex_2122, depress_2122, headache_2122, learning_2122, and others suggests that their influence on ADHD prediction is weaker and less consistent across children.

4.2 Deployment and Integration

The proposed AI-Driven ADHD Prediction and Analysis system has been effectively deployed to make the solution accessible and user-friendly in real-world settings. The proposed system has been deployed into both web and mobile application platforms for user convinience.

Web application(Streamlit)

The proposed ensemble stacking model has been deployed into Streamlit[59] for interactive web-based access and user-friendly activities.

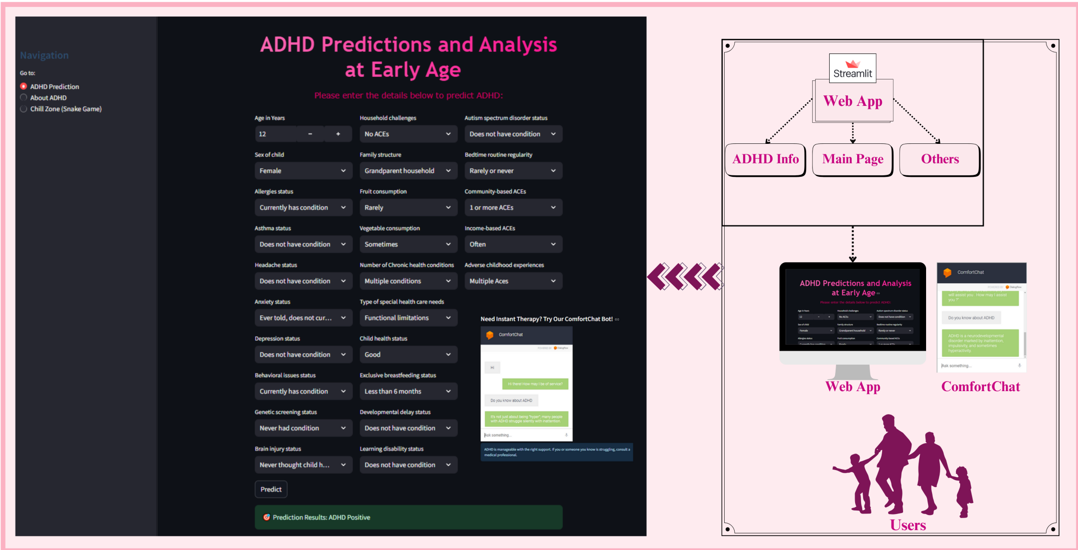

Figure 4.9: Interface of Webapp

Figure 4.9 , illustrates the home and prediction page of the developed web application. Moreover, it contains a navigation bar which includes main page, about ADHD page and others page. The chatbot is integrated both in main page and about ADHD page. Users are able to input 25 different questionarries and predict the ADHD outcome.

Mobile application(Gradio)

Apart from deployment in web application , the ensemble stacking model also has been deployed into mobile interface through Gradio[60].

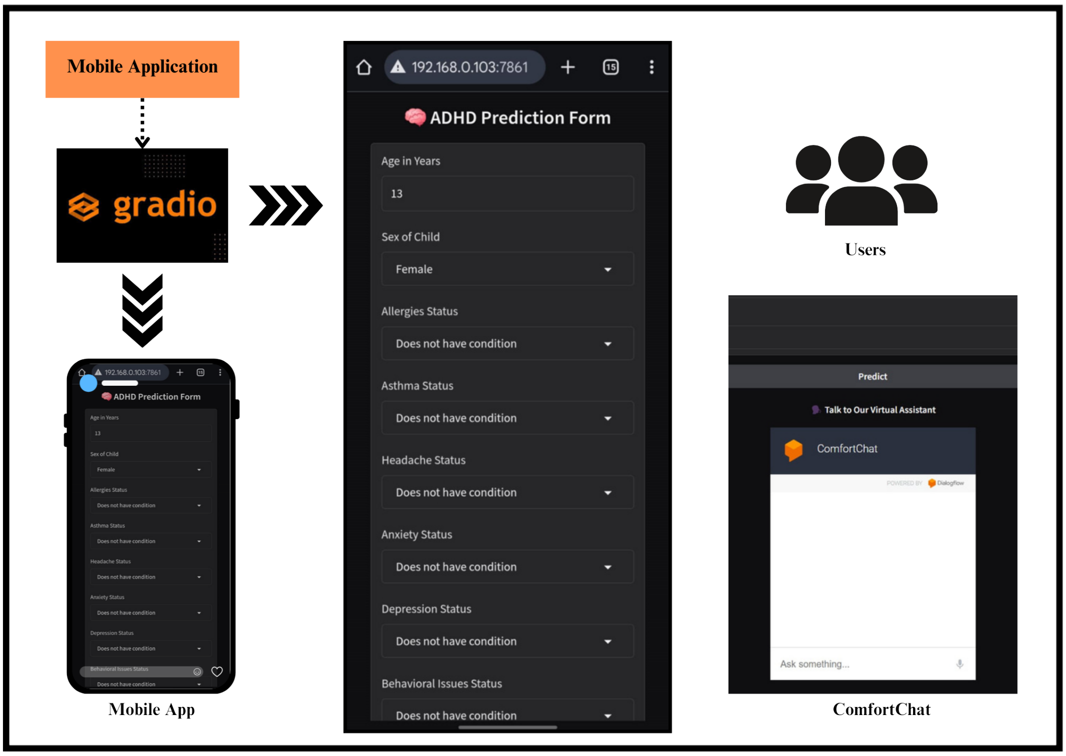

Figure 4.10: Interface of Mobileapp

The mobile application also kept user-friendly and interactive integrating the chatbot named ComfortChat. Figure 4.10 portrays the mobile app interface and external looks.

4.3 Discussion

This study introduces a novel and integrative approach for early ADHD detection in children by combining machine learning, large language models (LLMs), explainable AI (XAI), and a therapeutic chatbot. It addresses six core research questions, as outlined in the Introduction section.

For Research Question 1, Table 4.1 confirms that the proposed stacking ensemble model—built with Random Forest, LightGBM, and XGBoost as base learners and logistic regression as the meta-learner—achieved top performance (accuracy: 94.27%, precision: 87.00%, recall: 82.07%, F1: 84.30%). This clearly validates the strength of ML-based ADHD prediction. Comparison with related works and results from the generalization study (Section 4.0.2) further support the robustness of this approach.

For Research Question 2, interpretability was ensured using LIME, SHAP, permutation importance, and decision tree surrogates (Section 4.1). These XAI methods revealed that chronic illnesses and adverse childhood conditions significantly influence ADHD outcomes, thereby improving transparency and trust.

For Research Question 3, Table 4.7 highlights the effectiveness of BERT-based LLMs—especially DistilBERT—which achieved 93.45% accuracy and an AUC of 0.968 while maintaining the lowest runtime. Figure 4.6 illustrates the efficiency and consistency of transformer models in text-based ADHD analysis.

For Research Question 4, the Dialogflow-based chatbot 'ComfortChat' (Figure 3.13) was introduced to offer therapeutic aid, accessible through both web and mobile interfaces. It provides users—including caregivers and parents—with preliminary guidance and mental health support in a lightweight and user-friendly format.

Research Question 5 explored how socio-environmental factors, particularly Adverse Childhood Experiences (ACEs) and parental support, impact ADHD. Section 3.1.2, especially Subsection 3.1.2, reveals that ADHD prevalence rises dramatically with ACE exposure—from 8.8% (no ACEs) to 30.03% (with discrimination). (Figure 3.6) reinforce the role of environmental and household context in ADHD likelihood.

Finally, Research Question 6 validates the system's real-world applicability. Performance on unseen datasets (Tables 4.8 and 4.9) shows impressive generalization, with the blending ensemble reaching 94.98% accuracy on the 2022–2023 NSCH dataset—surpassing all prior results. Additionally, expert feedback from institutions like NIMH and Enam Medical College, as well as diverse stakeholders (including parents, clinicians, and AI researchers), confirmed both the system's validity and potential impact in practical deployment scenarios.

To the best of our knowledge, this is the first study to propose a comprehensive solution that combines predictive modeling and a therapeutic chatbot for ADHD support. The integration of generalization on unseen datasets and validation from domain experts enhances both credibility and real-world applicability. These components together form the core novelty of our proposed framework.

Chapter 5

Impacts of the Project

The proposed AI-driven ADHD prediction and intervention system demonstrates significant impacts across social, healthcare, educational, awareness, and environmental domains.

5.1 Societal, Healthcare, Educational, and Awareness Impact

Societal Impact: The system helps combat stigma and misinformation surrounding ADHD by enabling inclusive, judgment-free conversations through a virtual therapy chatbot. It promotes mental health education for all, regardless of age, gender, or background, while fostering community engagement by involving parents, caregivers, and educators in early detection. Its mobile/web accessibility ensures reach across socio-economic groups.

Healthcare Impact: This solution empowers users, caregivers, and professionals with a tool for early ADHD screening, supplementing clinical diagnosis. Its 24/7 accessibility, combined with transparent Explainable AI (via SHAP and LIME), enhances trust, reduces clinical dependency, and encourages early interventions for improved outcomes.

Educational Impact: Early identification supports better academic performance and behavior management. Teachers gain insights into student behavior, improving classroom dynamics. The chatbot also acts as an educational aid, offering ADHD management tips to schools and families.

Awareness Impact: The project raises public awareness by delivering accessible, science-based ADHD insights through LLMs and NLP. It enables self-assessment, highlights environmental and social risk factors (like ACEs), and encourages empathy, early support, and informed parenting.

5.2 Environmental and Sustainability Impact

This digital project indirectly contributes to environmental sustainability while strongly supporting social well-being. By replacing paper-based forms, minimizing clinical visits, and reducing resource-heavy infrastructures, it cuts down on carbon emissions and technological waste.

Its deployment on mobile and web platforms ensures efficient use of existing devices, aligning with green computing principles. From a broader perspective, the project supports multiple Sustainable Development Goals (SDGs)—including good health (SDG 3), reduced inequality (SDG 10), sustainable communities (SDG 11), and responsible consumption (SDG 12). It serves as a model for eco-conscious healthcare innovation by transforming traditional systems into efficient, scalable, and low-impact digital solutions.

Moreover, the reduced reliance on physical infrastructure helps conserve energy and space within medical facilities, promoting sustainable urban development. The system's ability to operate on lightweight hardware minimizes electronic waste generation. By eliminating the need for printed educational material, the platform preserves natural resources and supports paperless awareness campaigns. Its remote functionality reduces traffic congestion and carbon emissions tied to clinic commutes. Overall, the project advocates for an intersection of digital innovation and environmental responsibility in future healthcare solutions.

Chapter 6

Project Planning and Budget

A detailed project management plan is developed to successfully execute the project " AI-Driven ADHD Prediction and Analysis at Early Age: a Novel approach integrating Machine Learning, Explainable AI, LLMs, and Dialogflow with a Virtual therapy chatbot ". This plan provides adequate resource allocation, defined roles and responsibilities, timely execution of tasks, and quality assurance.

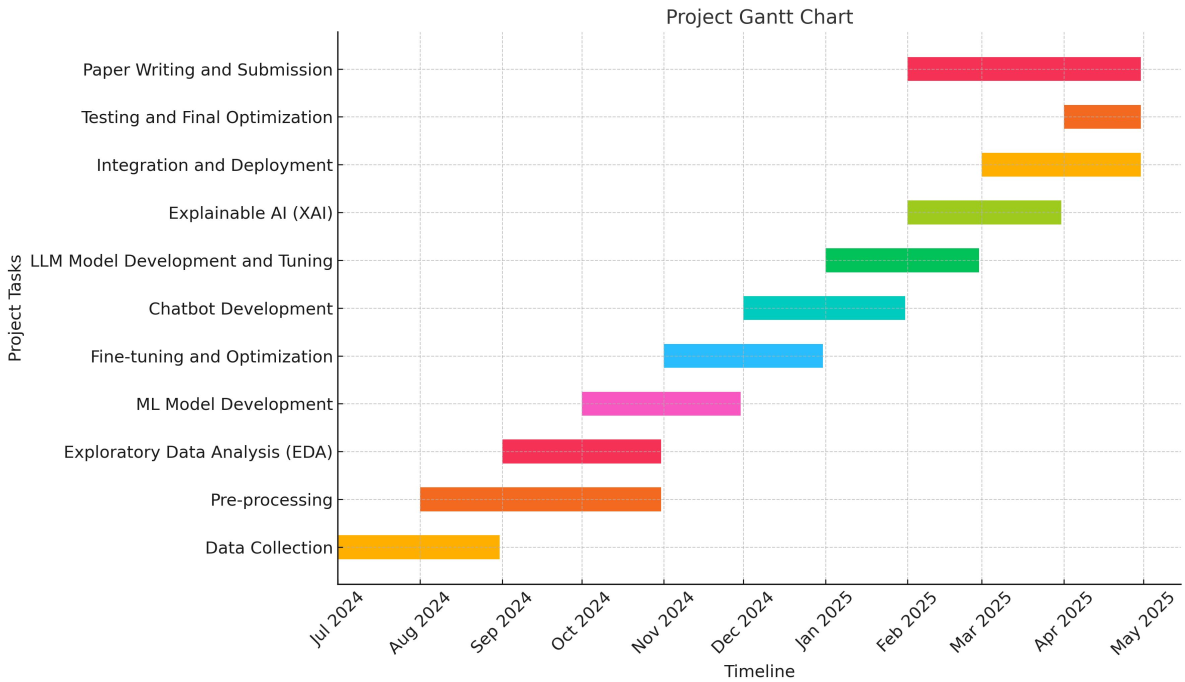

Figure 6.1: Gantt Chart For the Proposed Project (July 2024 - April 2025)

This Gantt chart in Fig. 6.1 outlines a detailed timeline for the ADHD prediction at early age project. Here's a breakdown of the project planning and task allocation based on the chart:

6.1 Project Planning

Project Timeline and Phases

1. **Data Collection (July 2024 – August 2024)**
   The first focus is gathering data from reputable sources that include indicators potentially related to ADHD symptoms.
2. **Pre-processing (August 2024 – October 2024)**
   This phase focuses on cleaning and preprocessing the collected data. Issues such as missing values, outliers, and inconsistencies will be addressed to ensure the dataset is ready for further analysis.
3. **Exploratory Data Analysis (EDA) (September 2024 – October 2024)**
   In-depth analysis will be conducted to identify patterns and key features that may help predict ADHD. This includes visualizations, correlation analysis, and feature selection for model building.
4. **ML Model Development (October 2024 – November 2024)**
   Machine learning models will be developed to detect ADHD symptoms in children using selected features. This includes training, validating, and testing different algorithms to find the most effective model.
5. **Fine-tuning and Optimization (November 2024 – December 2024)**
   The selected model will be fine-tuned to enhance its accuracy and efficiency through hyperparameter tuning and evaluation using performance metrics.
6. **Chatbot Development (December 2024 – January 2025)**
   A Dialogflow-based virtual therapy chatbot will be developed to provide ADHD-related resources and support. The chatbot will be trained, tested, and integrated with the prediction model for a seamless user experience.
7. **LLM Model Development and Tuning (January 2025 – February 2025)**
   A large language model (LLM) will be fine-tuned on ADHD-specific datasets to improve the chatbot's response quality, making interactions more informative and relevant.
8. **Explainable AI (XAI) (February 2025 – March 2025)**
   XAI techniques will be applied to interpret the model's predictions, increasing transparency and user trust by providing meaningful insights into how decisions are made.

9. **Integration and Deployment (March 2025 – April 2025)**
   All components—including the prediction model, chatbot, and LLM—will be integrated into an Android application, with full system deployment and functionality testing.
10. **Testing and Final Optimization (April 2025)**
    The final phase includes rigorous usability testing, bug fixes, and system optimization to ensure a robust, reliable, and user-friendly solution.
11. **Paper Writing and Submission (February 2025 – April 2025)**
    Project findings, methodology, and outcomes will be documented and prepared for submission to an academic journal. This includes writing, editing, formatting, and preparing visual aids such as figures and tables.

6.2 Budget

The estimated budget for the project is **BDT 500,000**, accounting for salaries, consultation fees, app development, and other necessary expenses.

Expense Item	**Cost (BDT)**	**Description**
Data Collection	30,000	Cost related to data collection
Researcher Salaries	300,000	Salary of 3 researchers @ 30,000 BDT/month for 10 months
Consultation Fees	100,000	Expert consultation for guidance, covering two consultants @ 50,000 BDT each
App Development Cost	50,000	The development cost for the app, virtual therapy chatbot, and system integration
Software and Tools	20,000	Software licenses and cloud computing resources
TOTAL	**500,000**	

Table 6.1: Estimated Project Budget

Budget Breakdown

This budget ensures the efficient allocation of resources across critical project components.

1. **Data Collection** (Total Cost: BDT 30,000)
   Covers expenses related to data collection, including surveys, tools, or acquir-

ing necessary datasets.
Percentage of Total Budget: 6%

2. **Researcher Salaries** (Total Cost: BDT 300,000)
Three researchers will work on this project for a duration of 10 months, each receiving a monthly compensation of 10,000 BDT. Their responsibilities will include data analysis, model development, and research paper preparation.
Percentage of Total Budget: 60%

3. **Consultation Fees** (Total Cost: BDT 100,000)
Two expert consultants, each receiving a fee of 50,000 BDT, will be engaged to provide strategic guidance, evaluate the model design, and offer technical or clinical insights on ADHD detection and therapy integration.
Percentage of Total Budget: 20%

4. **App Development Cost** (Total Cost: BDT 50,000)
This budget covers the development of an interactive app, including the integration of a virtual therapy chatbot and Dialogflow. This will ensure a user-friendly and secure platform.
Percentage of Total Budget: 10%

5. **Software and Tools** (Total Cost: BDT 20,000)
Funds have been allocated to purchase essential software licenses and secure cloud computing resources, including machine learning tools, data storage platforms, and APIs required for model development and deployment.
Percentage of Total Budget: 4%

Chapter 7

Complex Engineering Problems and Activities

7.1 Complex Engineering Problems (CEP)

Table 7.1: Table 0.1: A Sample of Complex Engineering Problem Attributes

Attributes	Addressing the Complex Engineering Problems (P) in the Project
P1: Depth of knowledge required (K3–K8)	The project requires knowledge of mathematical concepts in ML (K2), Python programming and ML libraries (K3), data preprocessing, feature engineering, Explainable AI (SHAP, LIME), deep learning (CNNs), LLMs (K4), Dialogflow integration (K5), web/chatbot development (K6), ADHD and its social context (K7), and research skills for academic writing (K8).
P2: Range of conflicting requirements	The system balances accuracy (94.27%), interpretability (XAI), efficiency (DistilBERT), usability (via chatbot), and deployment feasibility—often presenting conflicting trade-offs.
P3: Depth of analysis required	Significant preprocessing, feature selection, and class imbalance handling were required. EDA was performed on 800+ survey features, identifying ADHD-related patterns. Data cleaning (missing/outliers), training of 14 ML and 5 transformer models, metric analysis, validation on unseen data, and XAI interpretation (SHAP/LIME) were all integral.
P4: Familiarity of issues	The project addresses pediatric mental health—an uncommon domain in engineering. Challenges include limited standard datasets, integration of ACEs, cultural factors relevant to Bangladesh, and combining XAI with LLMs for healthcare support.
P5: Extent of applicable codes	There are no standard frameworks for ADHD classification with integrated XAI and chatbot therapy. Custom pipelines were developed for full-stack ML and conversational AI deployment across platforms.
P7: Interdependence	Multiple subsystems—data processing, modeling, XAI, Dialogflow, and chatbot—are interdependent. For instance, poor feature selection reduces model quality, which can cascade into irrelevant chatbot outputs and loss of user trust.

Table 7.1 demonstrates a sample complex engineering problem attribute.

7.2 Complex Engineering Activities (CEA)

Table 7.2: Table 0.2: A Sample of Complex Engineering Activity Attributes

Attributes	**Addressing the Complex Engineering Activities (A) in the Project**
A1: Range of resources	The project utilizes diverse resources, including CAHMI national health survey data, advanced ML/LLM models, cloud computing platforms, Dialogflow for chatbot integration, and mobile/web deployment technologies. Resources also include domain expert consultations, user interface tools, and psychological evaluation frameworks.
A3: Innovation	This project introduces a novel fusion of ADHD prediction through ML, transparency via SHAP/LIME, and personalized support through LLM-powered chatbot therapy. The chatbot provides educational and emotional guidance based on AI-driven insights. The solution employs innovative integration of state-of-the-art ML models, Explainable AI techniques, and LLMs to form a cohesive ADHD diagnostic and support system.
A4: Consequences to society/environment	The project addresses a serious public health concern—early ADHD detection—aiming to reduce long-term psychological distress, academic struggles, and societal costs. Its broad deployment potential across socio-economic contexts offers a scalable, accessible mental health support platform.
A5: Familiarity	The activities extend beyond previous experience by combining ML, XAI, LLMs, and therapeutic chatbot integration into a single system, something rarely addressed in traditional coursework or local healthcare solutions. This multi-layered pipeline requires constant adaptation and testing of unconventional engineering solutions.

Table 7.2 demonstrates a sample complex engineering activity attribute.

The attributes presented in Tables 7.1 and 7.2 collectively highlight the complexity and depth of the engineering challenges addressed in this project. From the need for multidisciplinary knowledge and advanced analytical techniques to managing conflicting requirements and system interdependence, the project meets the criteria of complex engineering problems. Simultaneously, the activities demonstrate the innovative use of resources, cross-functional collaboration, and societal relevance. The integration of AI, Explainable AI, and LLMs within a healthcare framework showcases not only technical innovation but also a strong alignment with sustainable and inclusive engineering practices. These combined factors affirm the project's status as a significant and comprehensive engineering endeavor.

Chapter 8

Conclusions

8.1 Summary

This system integrates machine learning, explainable AI, large language models, and a virtual therapy chatbot. This comprehensive approach seeks to ensure sound prediction and offers valuable insights and guidance for children, caregivers, and healthcare providers. In this system, machine learning models have been highly effective in diagnosing ADHD, with the Stacking ensemble model being highly accurate at 94.27% accuracy which achieves the state-of-art performance in CAHMI dataset. It establishes the potential of AI to enable diagnostic capabilities. Also, implementing explainable AI tools like LIME and SHAP allows transparency to the black box of AI and makes the system's predictions trustworthy. Developing Comfort Chat, a Dialogflow-based virtual therapist chatbot, adds a unique dimension to our work. It is a representation of dedication towards providing accessible, personalized guidance and care, transcending the dual challenges of stigma and limited access to specialist treatment. In contrast to earlier studies that concentrate only on prediction, our approach integrates early diagnosis with real-time therapeutic support, effectively closing the gap between identifying ADHD and initiating intervention.

8.2 Limitations

Despite its promising outcomes, this study has certain limitations. The reliance on survey-based data introduces potential biases and lacks neurophysiological, genetic, or medical records that could enhance ADHD diagnosis. Additionally, the absence of multimodal data, such as speech patterns, handwriting analysis, and eye-tracking, limits the model's ability to capture the full complexity of ADHD.

8.3 Future Improvement

Our future development plan is to broaden the scope of this work by systematically collecting and analyzing ADHD-related data from diverse regions across Bangladesh. This will enable the proposed system to be culturally contextualized, ensuring its relevance and effectiveness in local communities. In addition to expanding the dataset, we aim to incorporate multimodal inputs—such as EEG signals, MRI scans, behavioral videos, and audio recordings—to enhance the depth and accuracy of ADHD prediction. This multimodal integration could significantly improve diagnostic capabilities by capturing a wider spectrum of neurodevelopmental indicators. The application will be further enhanced with user-centered design principles and deployed through a dedicated web platform, ensuring free, user-friendly, and scalable access for clinicians, educators, and families. Moreover, future collaborations with healthcare institutions, academic researchers, and government agencies are planned to support nationwide deployment, ethical oversight, and long-term impact evaluation. This broader vision positions our system as a transformative tool in early ADHD detection and mental health intervention.

Bibliography

[1] F. Svenaeus. Diagnosing mental disorders and saving the normal. *Medicine, Health Care and Philosophy*, 17:241–244, 2014.

[2] S. Cortese. The neurobiology and genetics of attention-deficit/hyperactivity disorder (adhd): what every clinician should know. *European Journal of Paediatric Neurology*, 16(5):422–433, 2012. Epub 2012 Feb 2.

[3] G. V. Polanczyk, G. A. Salum, L. S. Sugaya, A. Caye, and L. A. Rohde. Annual research review: a meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry*, 56:345–365, 2015.

[4] J. L. Barican, D. Yung, C. Schwartz, Y. Zheng, K. Georgiades, and C. Waddell. Prevalence of childhood mental disorders in high-income countries: a systematic review and meta-analysis to inform policymaking. *Evidence-Based Mental Health*, 25:36–44, 2022.

[5] E. J. Sonuga-Barke, J. Koerting, E. Smith, D. C. McCann, and M. Thompson. Early detection and intervention for attention-deficit/hyperactivity disorder. *Expert Review of Neurotherapeutics*, 11(4):557–563, 2011.

[6] O. Zahmacioğlu and E. Z. Kilic. Early diagnosis and treatment of adhd are important for a secure transition to adolescence. *Anatolian Journal of Psychiatry / Anadolu Psikiyatri Dergisi*, 18(1), 2017.

[7] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*. American Psychiatric Publishing, 2013.

[8] C. A. Boyle, S. Boulet, L. A. Schieve, R. A. Cohen, S. J. Blumberg, M. Yeargin-Allsopp, S. Visser, and M. D. Kogan. Trends in the prevalence of developmental disabilities in us children, 1997-2008. *Pediatrics*, 127(6):1034–1042, 2011.

[9] Centers for Disease Control and Prevention. Prevalence of diagnosis and medication treatment for adhd - united states, 2003. *MMWR Morbidity and Mortality Weekly Report*, 54(34):842–847, 2005.

[10] S. N. Visser, M. L. Danielson, R. H. Bitsko, J. R. Holbrook, M. D. Kogan, R. M. Ghandour, R. Perou, and S. J. Blumberg. Trends in the parent-report of health care provider-diagnosed and medicated attention-deficit/hyperactivity disorder: United states, 2003–2011. *Journal of the American Academy of Child & Adolescent Psychiatry*, 53(1):34–46.e32, 2014.

[11] B. Zablotsky, L. I. Black, M. J. Maenner, L. A. Schieve, M. L. Danielson, R. H. Bitsko, S. J. Blumberg, M. D. Kogan, and C. A. Boyle. Prevalence and trends of developmental disabilities among children in the us, 2009-2017. *Pediatrics*, 144(4):e20190811, 2019.

[12] Yang S. Liu et al. Early identification of children with attention-deficit/hyperactivity disorder (adhd). *PLOS Digital Health*, 3(11):e0000620, 2024.

[13] Susan Young et al. Failure of healthcare provision for attention-deficit/hyperactivity disorder in the united kingdom: A consensus statement. *Frontiers in Psychiatry*, 12:649399, 2021.

[14] S. Grazioli, A. Crippa, E. Rosi, A. Candelieri, S. B. Ceccarelli, M. Mauri, M. Manzoni, V. Mauri, S. Trabattoni, M. Molteni, P. Colombo, and M. Nobile. Exploring telediagnostic procedures in child neuropsychiatry: addressing adhd diagnosis and autism symptoms through supervised machine learning. *European Child and Adolescent Psychiatry*, 33(1):139–149, January 2024. Epub 2023 Jan 25.

[15] Md. Maniruzzaman, J. Shin, and Md. Al Mehedi Hasan. Predicting children with adhd using behavioral activity: A machine learning analysis. *Applied Sciences*, 12(5):2737, 2022.

[16] W. Das and S. Khanna. A robust machine learning based framework for the automated detection of adhd using pupillometric biomarkers and time series analysis. *Scientific Reports*, 11:16370, 2021.

[17] H. Qin, L. Zhang, J. Wang, et al. Interpretable machine learning approaches for children's adhd detection using clinical assessment data: An online web application deployment. *BMC Psychiatry*, 25:139, 2025.

[18] S. Kim, H. K. Lee, and K. Lee. Can the mmpi predict adult adhd? an approach using machine learning methods. *Diagnostics*, 11(976), 2021.

[19] J. Shin, M. Maniruzzaman, Y. Uchida, M. A. M. Hasan, A. Megumi, and A. Yasumura. Handwriting-based adhd detection for children having asd using machine learning approaches. *IEEE Access*, 11:84974–84984, 2023.

[20] M. Duda, R. Ma, N. Haber, and D. P. Wall. Use of machine learning for the behavioral distinction of autism and adhd. *Translational Psychiatry*, 6(2):e732, 2016.

[21] Nora Fink. A high-accuracy supervised machine learning approach for adhd classification using eeg signals. `https://ssrn.com/abstract=5146511`, 2025. Available at SSRN: `https://ssrn.com/abstract=5146511` or `http://dx.doi.org/10.2139/ssrn.5146511`.

[22] Catherine Joy, Thomas S. George, Albert A. Rajan, and M. S. P. Subathra. Detection of adhd from eeg signals using different entropy measures and ann. *Clinical EEG and Neuroscience*, 53(1):12–23, 2022.

[23] Md. Maniruzzaman, Koki Hirooka, Yoichi Tomioka, Md. Al Mehedi Hasan, Yong Seok Hwang, Akiko Megumi, Akira Yasumura, and Jungpil Shin. Machine learning-based adhd detection from fnirs signal during reverse stroop tasks. *IEEE Access*, 12:82984–82995, 2024.

[24] C. Nash, R. Nair, and S. M. Naqvi. Machine learning in adhd and depression mental health diagnosis: A survey. *IEEE Access*, 11:86297–86317, 2023.

[25] M. Garcia-Argibay, Y. Zhang-James, S. Cortese, et al. Predicting childhood and adolescent attention-deficit/hyperactivity disorder onset: a nationwide deep learning approach. *Molecular Psychiatry*, 28:1232–1239, 2023.

[26] P. Mikolas, A. Vahid, F. Bernardoni, et al. Training a machine learning classifier to identify adhd based on real-world clinical data from medical records. *Scientific Reports*, 12:12934, 2022.

[27] S. Khanna and W. Das. A novel application for the efficient and accessible diagnosis of adhd using machine learning (extended abstract). In *2020 IEEE/ITU International Conference on Artificial Intelligence for Good (AI4G)*, pages 51–54, 2020.

[28] W. P. Kim, H. J. Kim, S. P. Pack, J. H. Lim, C. H. Cho, and H. J. Lee. Machine learning-based prediction of attention-deficit/hyperactivity disorder and sleep problems with wearable data in children. *JAMA Network Open*, 6(3):e233502, 2023.

[29] Y. Zhang-James, Q. Chen, R. Kuja-Halkola, P. Lichtenstein, H. Larsson, and S. V. Faraone. Machine-learning prediction of comorbid substance use disorders in adhd youth using swedish registry data. *Journal of Child Psychology and Psychiatry*, 61:1370–1379, 2020.

[30] Child and Adolescent Health Measurement Initiative (CAHMI). 2021-2022 national survey of children's health (nsch) indicator dataset. Data Resource

Center for Child and Adolescent Health, supported by HRSA, MCHB, 2025. Retrieved February 24, 2025, from www.childhealthdata.org.

[31] Johns Hopkins Bloomberg School of Public Health. Child and adolescent health measurement initiative (cahmi). https://www.jhsph.edu/cahmi, 2023.

[32] X. Chu, I. F. Ilyas, S. Krishnan, and J. Wang. Data cleaning: Overview and emerging challenges. In *Proceedings of the 2016 International Conference on Management of Data*, pages 2201–2206. ACM, June 2016.

[33] W. McKinney et al. pandas: Powerful data analysis toolkit. https://pandas.pydata.org/, 2022. Accessed: August 6, 2022.

[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[35] M. M. Ahsan, M. P. Mahmud, P. K. Saha, K. D. Gupta, and Z. Siddique. Effect of data scaling methods on machine learning algorithms and model performance. *Technologies*, 9(3):52, 2021.

[36] S. Sperandei. Understanding logistic regression analysis. *Biochemia Medica*, 24(1):12–18, 2014.

[37] Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc., Sebastopol, CA, 2017.

[38] L. E. Peterson. K-nearest neighbor. *Scholarpedia*, 4(2):1883, 2009.

[39] S. Chen, G. I. Webb, L. Liu, and X. Ma. A novel selective naïve bayes algorithm. *Knowledge-Based Systems*, 192:105361, 2020.

[40] M. Awad and R. Khanna. *Support Vector Machines for Classification*. A Press, Berkeley, CA, 2015.

[41] M. Belgiu and L. Drăguţ. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:24–31, 2016.

[42] Candice Bentéjac, Anna Csörgő, and Gonzalo Martínez-Muñoz. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54:1937–1967, 2021.

[43] B. Charbuty and A. Abdulazeez. Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(1):20–28, March 2021.

[44] A. K. Singh. Breast cancer classification using ml on wdbc. In *Lecture Notes in Electrical Engineering*, pages 609–619. January 2023.

[45] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

[46] J. Zhang, D. Mucs, U. Norinder, and F. Svensson. Lightgbm: An effective and scalable algorithm for prediction of chemical toxicity–application to the tox21 and mutagenicity data sets. *Journal of Chemical Information and Modeling*, 59(10):4150–4158, 2019.

[47] J. T. Hancock and T. M. Khoshgoftaar. Catboost for big data: An interdisciplinary review. *Journal of Big Data*, 7:94, 2020.

[48] B. Pavlyshenko. Using stacking approaches for machine learning models. In *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, pages 255–258, 2018.

[49] Nicos Isaak and Loizos Michael. Blending nlp and machine learning for the development of winograd schemas. 03 2021.

[50] R. U. Mhapsekar, N. O'Shea, S. Davy, and L. Abraham. Hybrid blended deep learning approach for milk quality analysis. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(3):2253–2268, June 2024.

[51] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint*, arXiv:1810.04805, 2018.

[52] X. Jiao, Y. Yin, L. Shang, X. Jiang, X. Chen, L. Li, F. Wang, and Q. Liu. Tinybert: Distilling bert for natural language understanding. *arXiv preprint*, arXiv:1909.10351, 2019.

[53] Kexin Huang, Jaan Altosaar, and Rajesh Ranganath. Clinicalbert: Modeling clinical notes and predicting hospital readmission. *arXiv preprint*, arXiv:1904.05342, 2019.

[54] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang. Biobert: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, 2020.

[55] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint*, arXiv:1909.11942, 2020.

[56] A. F. Muhammad, D. Susanto, A. Alimudin, F. Adila, M. H. Assidiqi, and S. Nabhan. Developing english conversation chatbot using dialogflow. In *2020*

International Electronics Symposium (IES), pages 468–475. IEEE, September 2020.

[57] R. Meyes, M. Lu, C. W. de Puiseau, and T. Meisen. Ablation studies in artificial neural networks. *arXiv preprint*, arXiv:1901.08644, 2019.

[58] Johns Hopkins Bloomberg School of Public Health. Cahmi - child and adolescent health measurement initiative. `https://publichealth.jhu.edu/departments/population-family-and-reproductive-health/research-and-practice/life-course-framework/child-and-adolescent-health/cahmi-child-and-adolescent-health-measurement-initiative`, 2025. Accessed: February 2025.

[59] Streamlit Inc. Streamlit — the fastest way to build and share data apps. `https://streamlit.io`, 2023. Accessed: February 2025.

[60] Gradio Labs. Gradio: Build machine learning web apps in python. `https://www.gradio.app`, 2023. Accessed: February 2025.