Development of Machine Learning Models for Prediction of Psychosocial Dysfunction in Children and Adolescents using the Pediatric Symptom Checklist

By

Mahnoor Hasan

(Registration No: 00000401153)

Department of Sciences

School of Interdisciplinary Engineering & Sciences

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(May 2024)

Development of Machine Learning Models for Prediction of Psychosocial Dysfunction in Children and Adolescents using the Pediatric Symptom Checklist

By

Mahnoor Hasan

(Registration No: 00000401153)

A thesis submitted to the National University of Sciences and Technology, Islamabad,

in partial fulfillment of the requirements for the degree of

Masters of Science in
Bioinformatics

Supervisor: Dr. Zamir Hussain

School of Interdisciplinary Engineering & Sciences

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(May 2024)

14/5/24

Annex A to NUST Letter No.
0972/102/Exams/Thesis-Cert
dated 23 Dec 16.

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by Ms. Mahnoor Hasan Registration No. 00000401153 of SINES has been vetted by undersigned, found complete in all aspects as per NUST Statutes/Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature with stamp: Associate Professor SINES - NUST, Sector H-12 Islamabad

Name of Supervisor: Dr. Zamir Hussain

Date: 10-05-2024

Signature of HoD with stamp: Dr. Fouzia Malik HoD Sciences Professor SINES NUST, Sector H-12 Islamabad

Date: 15-5-2024

Countersign by

Signature (Dean/Principal): Dr. SYED IRTIZA ALI SHAH Principal & Dean SINES - NUST, Sector H-12 Islamabad

Date: 16 MAY 2024

CERTIFICATE OF APPROVAL

This is to certify that the research work presented in this thesis, entitled "Development of Machine Learning Models for Prediction of Psychosocial Dysfunction in Children and Adolescents using the Pediatric Symptom Checklist" was conducted by Ms. Mahnoor Hasan under the supervision of Dr. Zamir Hussain.

No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the SINES in partial fulfillment of the requirements for the degree of Master of Science in Field of Bioinformatics.

Department of Science, SINES, National University of Sciences and Technology, Islamabad.

Student Name: Mahnoor Hasan **Signature:** ________

Examination Committee:

a) **Examiner 1:** Dr. Zartasha Mustansar **Signature:** ________
HOD, Associate Professor, SINES

b) **Examiner 2:** Dr. Masood Ur Rehman Kayani **Signature:** ________
Assistant Professor, SINES

c) **External Examiner:** Dr. Qurrat Ulain Hamdan **Signature:** Qurrat Ulain
Assistant Professor, Institute of Psychiatry, Rawalpindi Medical University

Supervisor: Dr. Zamir Hussain **Signature:** ________

Dean: **Signature:** ________

Dr. SYED IRTIZA ALI SHAH
Principal & Dean
SINES - NUST, Sector H-12
Islamabad

AUTHOR'S DECLARATION

I Mahnoor Hasan hereby state that my MS thesis titled "Development of Machine Learning Models for Prediction of Psychosocial Dysfunction in Children and Adolescents using the Pediatric Symptom Checklist" is my own work and has not been submitted previously by me for taking any degree from National University of Sciences and Technology, Islamabad or anywhere else in the country/ world.

At any time if my statement is found to be incorrect even after I graduate, the university has the right to withdraw my MS degree.

Name of Student: Mahnoor Hasan

Date: 03-04-2024

PLAGIARISM UNDERTAKING

I solemnly declare that research work presented in the thesis titled "Development of Machine Learning Models for Prediction of Psychosocial Dysfunction in Children and Adolescents using the Pediatric Symptom Checklist" is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and National University of Sciences and Technology (NUST), Islamabad towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the University reserves the rights to withdraw/revoke my MS degree and that HEC and NUST, Islamabad has the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.

Student Signature: ____________

Name: Mahnoor Hasan

DEDICATION

I dedicate this thesis to my beloved parents, and my dear friends and colleagues

Ms. Hafsa Amjad and Ms. Irza Mahmood.

ACKNOWLEDGEMENTS

My profound gratitude is directed towards Almighty Allah for the divine blessings that sustained me throughout the demanding yet fulfilling pursuit of this research. His unwavering support empowered me to navigate challenges and ultimately achieve my desired goals, resulting in the successful completion of this project.

I extend my deepest gratefulness to my esteemed supervisor, Dr. Zamir Hussin, whose support, insightful guidance, and profound knowledge were helpful in the successful completion of this research. His constant motivation, enthusiasm, and patience were indispensable throughout the entire process. Undoubtedly, his dedication to research serves as a perpetual source of inspiration for me.

Furthermore, I acknowledge the pivotal role played by the School of Interdisciplinary Engineering and Sciences (SINES), NUST, in providing the necessary research infrastructure and enabling environment for this project. I am deeply indebted to the members of my Guidance and Examination Committee (GEC), Dr. Zartasha Mustansar, Dr. Masood Ur Rehman Kayani, and Dr. Qurrat Ulain Hamdan (External), whose expert guidance, critiques, and support proved invaluable throughout the research process.

I extend my special appreciation to the team at SehatYab, who graciously provided the data that made this research possible, and to my research fellow, Hafsa Amjad, whose assistance at key moments significantly contributed to the progress of this research.

TABLE OF CONTENTS

LIST OF TABLES

Page No.

LIST OF FIGURES

Page No.

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
ML	Machine Learning
PSC	Pediatric Symptom Checklist
R	Pearson Correlation Coefficient
χ^2	Chi-square Statistic
C	Contingency Coefficient
RFE	Recursive Feature Elimination
DT	Decision Tree
RF	Random Forest
XGB	Extreme Gradient Boosting or XGBoost
SVM	Support Vector Machine
MLP	Multilayer Perceptron
LR	Logistic Regression

ABSTRACT

Mental health issues, particularly early-onset psychopathology, are a neglected domain of public health. In low middle-income countries like Pakistan, the management of these problems is hindered by stigma, limited mental healthcare resources, and protracted nature of consultations. The access and convenience of mental healthcare can be improved by leveraging Artificial Intelligence (AI) for the development of data-driven decision support systems. This study aims to develop Machine Learning (ML) based predictive models for psychosocial dysfunction in Pakistani children and adolescents. The Pediatric Symptom Checklist (PSC) is used to collect data of 2,372 individuals. To the best of our knowledge, this study is the first to use data of Pakistani pediatric populations collected using the PSC. Feature selection methods reveal that items related to attention issues, hopelessness, sadness, and irritability are the most significant predictors of psychosocial dysfunction for both age groups. Exclusively for children, externalizing problems like disobedience are also significant. Six ML algorithms viz. Decision Tree, Random Forest, XGBoost, Support Vector Machine, Multilayer Perceptron, and Logistic Regression are selected to develop the predictive models. Among these, Logistic Regression provides the best results with accuracies of 0.98 and 0.99 for children and adolescents, respectively. For children, it is observed that exclusion of item 23 of the PSC (Wants to be with you more than before) improves model performance. Our encouraging findings suggest that the proposed models have the potential to be deployed in healthcare facilities and educational institutes for early and accurate detection of pediatric psychosocial dysfunction.

Keywords: Psychosocial Dysfunction, Child and Adolescent Mental Health, Pediatric Symptom Checklist, Machine Learning, Artificial Intelligence.

1. INTRODUCTION

The formative years of childhood and adolescence are a period of significant social and cognitive development, laying the foundation of an individual's future trajectory in life [1]. Abnormalities that arise during these critical phases can negatively impact a young individual's emotions, intellect, and relationships [2]. Moreover, early onset psychopathology can persist and intensify in adulthood, leading to severe consequences like suicidality, addiction problems, and criminal behavior [3]. Globally, around 10-20% of children and adolescents suffer from mental health issues, where adolescents are more susceptible [4], [5]. Presently, the restrictive lifestyles of COVID-19 lockdowns and the devastation caused by the pandemic have also induced a significant psychological burden upon children due to an increase in stress, anxiety, irritation, and inattention [6]. Despite the escalating prevalence of pediatric mental health concerns, childhood and adolescent-onset psychopathology is a neglected domain within the realm of public health, especially in low middle-income (LMIC) developing countries like Pakistan.

1.1 Pediatric Mental Health in Pakistan

As of 2020, it is estimated that around 15 million Pakistanis suffer from some form of mental illness [7]. To the best of our knowledge, such empirical statistics for children and adolescents are not documented on a national level even though nearly half of the population of Pakistan is under the age of 18 [8]. However, surveys and researches carried upon limited sample sizes indicate that the burden of early-onset mental health issues exists. According to a survey of 1,124 adolescents in Rawalpindi, 17.2% and 21.4% were suspected of having depression and anxiety, respectively [9]. In Karachi, a survey of 1470 participants aged 11 to 17 years revealed that around 20% exhibit severe emotional and behavioral problems, while another survey of 640 adolescents estimates the prevalence of abnormal social and emotional behavior at 34% [10], [11]. Regardless of these reported estimates, resources for managing pediatric mental health problems do not emulate their severity.

1.1.1 *Challenges of Pediatric Mental Healthcare*

According to the Mental Health Atlas published by World Health Organization (WHO) in 2017, only 1% of the 3,729 outpatient mental health facilities in Pakistan specialize in child and adolescent psychiatry [12]. There are only around 400 trained psychiatrists, most of whom practice in urban areas [7]. In addition to scarce resources, the effective management of mental health issues is hampered by inadequate public awareness, societal stigma, and fallacious spiritual beliefs [13]. The associated ignorance and stigma ultimately lead to these issues remaining undetected, as clear signs are often depreciated as transient bursts of emotions. Limitations of mental health consultations also exacerbate the situation. The screening of mental health issues involves questionnaires and checklists in addition to one-on-one interviews. These screening tools are often extensive, with 30 to 100 questions per tool [14] (Table 1.1). On average, their required time of administration ranges from 30 to 60 minutes, which is quite long for both patients and mental healthcare professionals. In addition to the protracted nature of mental health consultations, their financial strain also discourages people from seeking help.

Table 1.1: Number of items in common mental health screening tools.

No.	Assessment	Mental Health Issue	Number of Items
1	Patient Health Questionnaire (PHQ) Screeners	Anxiety, Depression, Eating Disorders	83
2	Pediatric Symptom Checklist (PSC)	Psychosocial Dysfunction, Attention, Internalizing, and Externalizing Problems	35
3	Disruptive Behavior Rating Scale	Attention-Deficit/Hyperactivity Disorder	45
4	Vanderbilt Assessment Scales	Attention-Deficit/Hyperactivity Disorder	25 to 55
5	Screen for Child Anxiety Related Disorders (SCARED)	Anxiety Disorders	41
6	Spence Children’s Anxiety Scale (SCAS)	Anxiety Disorders	35 to 45
7	Child Behavior Checklist (CBCL)	Pervasive Developmental Disorder, Anxiety, Attention-Deficit/Hyperactivity Disorder	100-160
8	Swanson, Nolan, and Pelham (SNAP-IV) Scale	Conduct Disorders, Anxiety, Attention-Deficit/Hyperactivity Disorder	90

Another issue with current screening and diagnostic tools is their lack of standardization and generalizability. Most "universal" tools are developed and optimized according to European or American populations. For instance, the Pediatric Symptom Checklist was developed by American medical professionals and first validated on American cohorts [15]. Since then, it has been validated and optimized for a few African and Asian communities [16]. Unfortunately, these modifications have not included Pakistani populations. Sociodemographic factors of study populations are reported as key influences in the development of early-onset psychopathology [17]. Therefore, the standardization and optimization of screening tools with emphasis on Pakistani pediatric populations is necessary to ensure their efficacy.

In light of these observations, it is clear that the accessibility and efficacy of mental health care services warrants improvement. This can be accomplished by leveraging Artificial Intelligence (AI) for the development of smart data-driven clinical decision support systems. These smart tools facilitate healthcare professionals in evidence-based decision-making, thereby alleviating their workload and optimizing patient care [18]. Additionally, the adaptability and convenience of such user-friendly systems can also increase patient outcome. The current extent and scope of integration of AI in mental healthcare is thoroughly discussed in Chapter 2.

1.2 Psychosocial Dysfunction

Psychosocial dysfunction broadly encompasses irregular psychological and social behaviors like aggression, anxiety, depression, and attention issues [19], [20]. While psychosocial dysfunction itself does not indicate a specific diagnosis, it is documented as an indicator or precursor of personality and depressive disorders [21], [22], [23]. It may serve as a "starting point" to determine whether a child or adolescent requires professional help. This makes it an excellent choice as the target variable for ML-based screening models.

1.3 The Pediatric Symptom Checklist

The Pediatric Symptom Checklist (PSC) is a screening questionnaire for overall psychosocial dysfunction or psychological impairment in children and adolescents aged 3 to 17 years [15]. It consists of 35 questions or items which analyze the emotional, behavioral, and social aspects of a young individual's personality. The PSC can also be used to screen for specific problems as the questionnaire is divided into 3 subscales; attention, internalizing, and externalizing. As the names indicate, these subscales assess the risks of attention problems like attention-deficit and disorganization, internalizing problems like anxiety and depression, and externalizing problems like hyperactivity and aggression [24], [25]. All 35 items of the PSC and the items that constitute the 3 subscales are defined in Table 1.2.

Table 1.2: The Pediatric Symptom Checklist.

No.	Item	Question
1	PSC1	Complains of aches/pains
2	PSC2	Spends more time alone
3	PSC3	Tires easily, has little energy
4	PSC4	Fidgety, unable to sit
5	PSC5	Has trouble with a teacher
6	PSC6	Less interested in school
7	PSC7	Acts as if driven by a motor
8	PSC8	Daydreams too much
9	PSC9	Distracted easily
10	PSC10	Is afraid of new situations
11	PSC11	Feels sad, unhappy
12	PSC12	Is irritable, angry
13	PSC13	Feels hopeless
14	PSC14	Has trouble concentrating
15	PSC15	Less interest in friends
16	PSC16	Fights with others
17	PSC17	Absent from school
18	PSC18	School grades dropping
19	PSC19	Is down on him or herself
20	PSC20	Visits doctor with doctor finding nothing wrong
21	PSC21	Has trouble sleeping
22	PSC22	Worries a lot
23	PSC23	Wants to be with you more than before
24	PSC24	Feels he or she is bad
25	PSC25	Takes unnecessary risks
26	PSC26	Gets hurt frequently
27	PSC27	Seems to be having less fun
28	PSC28	Acts younger than children his or her age
29	PSC29	Does not listen to rules
30	PSC30	Does not show feelings
31	PSC31	Does not understand other people's feelings
32	PSC32	Teases others
33	PSC33	Blames others for his or her troubles
34	PSC34	Takes things that do not belong to him or her
35	PSC35	Refuses to share

Yellow Shading: Items of the Attention Subscale, Blue Shading: Items of the Internalizing Subscale, Green Shading: Items of the Externalizing Subscale

The PSC is a free and easy to interpret tool with excellent psychometric properties, making it suitable and favorable for the preliminary screening of pediatric psychosocial problems in children and adolescents [26]. In the context of ML and predictive modelling, the final result of the PSC has been used as a predictive feature for modeling the patterns of attrition in weight management programs for children [27]. However, to the best of our knowledge, the individual items of the PSC have not been analyzed using ML techniques or utilized for the development of smart screening processes thus far.

1.4 Problem Statement

Childhood and adolescence-onset mental health issues are a neglected area of public health concern due to societal stigma, lack of awareness, and inadequate mental healthcare resources. Moreover, existing intervention strategies are time-consuming and costly. These limitations in mental healthcare further exacerbate the escalating prevalence of early onset mental health issues. Failure to detect mental health issues that arise during childhood and adolescence can lead to serious consequences like addiction, suicidality, and personality disorders in adulthood. Therefore, the access, convenience, and efficacy of mental healthcare strategies necessitates improvement.

The enhancement of existing screening tools for mental health issues, specifically psychosocial dysfunction, is proposed as a possible solution to combat the challenges of mental healthcare. This can be accomplished by using ML algorithms for the development of smart data-driven decision support systems. These can aid mental healthcare professionals during the screening process by making rapid informed decisions based on local data. Additionally, the adaptability of such cost-friendly tools can increase their popularity and access among the general public, leading to an elevated patient outcome.

1.5 Objectives

This study aims to achieve the following objectives:

- Descriptive analysis of psychosocial dysfunction in Pakistani pediatric population with respect to their age and location.

- Identification of significant features for the prediction of psychosocial dysfunction in children and adolescents.
- Development of ML-based predictive models for the detection of psychosocial dysfunction that are optimized for Pakistani children and adolescents.

1.6 Relevance to national needs

In light of the statistics of mental health issues and limitations of mental healthcare described in Section 1.1, it is evident that research and development in the sector of mental health is required to accomplish the following:

- Development of rapid and cost-friendly screening processes for mental health issues to improve the accessibility and adaptability of mental healthcare.
- Provision of assistance to mental healthcare professionals to alleviate their workload and improve patient outcome.
- Optimization of mental healthcare according to Pakistani pediatric populations.

As a result of this study, the provision of mental health care could become more accessible for the general public via an automated screening tool. This will lead to faster detection of psychological problems among children and adolescents and promote their timely intervention and management. Ultimately, this will contribute to the Sustainable Development Goal (SDG) 3 - *Good Health and Well-being*. Moreover, the development of a quick screening tool will provide mental health care professionals with a smart decision support system, which will alleviate the burden of preliminary psychological consultations and simplify the process. This can ultimately achieve SDG 9 – *Industry, Innovation, and Infrastructure* by enhancing and upgrading the traditional healthcare approaches.

By using local data, this study will develop data-driven predictive models that are optimized according to Pakistani pediatric populations. Moreover, the availability and deployment of a smart processing tool is not limited to urban communities only. One of the benefits of such decision support systems is that they can also be implemented in rural and more remote locations. These potential outcomes can contribute to reducing inequalities in healthcare, thereby achieving SDG 10 – *Reduced Inequalities*.

1.7 Thesis Structure

This dissertation follows a comprehensive structure to successfully accomplish the aforementioned objectives in Section 1.5. Chapter 2 describes the literature review conducted to analyze the scope of research regarding AI-driven prediction of mental health issues in young populations and identify potential research gaps. Chapter 3 details the research methodology of the study, followed by Chapter 4 which discusses the obtained results in relation to published literature. Lastly, Chapter 5 concludes the dissertation with a summary of the study, acknowledged limitations, and recommendations to address them.

2. LITERATURE REVIEW

In the past decade, studies employing ML to enhance mental health screening and diagnostic procedures have emerged as a prominent domain of interdisciplinary research between computer science and healthcare [28]. ML is a subfield of AI inspired by the human brain's adaptive learning capabilities. It focuses on computational systems that can autonomously acquire knowledge and enhance their performance over time [29].

Within ML, two types of learning methods exist; supervised and unsupervised. Supervised ML means that the data being used to train the algorithm is already labelled with the target classes. In essence, the algorithm is taught or *supervised* to make decisions by explicitly telling it the corresponding decisions for each instance in the data [30]. In unsupervised ML, the data is not labelled. By analyzing the underlying patterns of the data, the algorithm itself designates labels to each instance [31]. In the context of ML-based prediction of mental health issues, or the use for ML in healthcare in general, supervised ML methods are more suitable as decisions regarding the health of an individual can and should only be made by a healthcare professional. ML algorithms should only be used to assist in this process.

Emotional and behavioral data collected via screening questionnaires, diagnostic interviews, medical histories, and general surveys can be analyzed using ML algorithms for the development of efficient predictive processes. The amalgamation of various types and large amounts of data for the screening of mental health issues makes the process more precise and generalizable, and the computational abilities of ML algorithms can analyze such vast data and obtain satisfactory results expeditiously. In addition to the prediction of potential mental health issues, ML algorithms can also be used to elucidate the most significant risk factors that increase the development or susceptibility of such issues. This type of analysis can inform specific psychological concerns that require special intervention.

The following sections explain current literature focusing on the detection of mental health issues in young individuals. The literature is categorized according to the type of mental health issue being detected.

2.1 ML for General Emotional and Behavioral Problems

2.1.1 Subjective Well-Being

Zhang et al. analyzed the state of Subjective Well-being (SWB) of 10,518 adolescents studying in the first 2 years of a medical degree in a Chinese university [32]. Gradient Boosting (GB) classifier was trained and tested upon clinical, socioeconomic, and environmental data. In addition to the prediction of SWB, associated risk factors were also examined. After hyperparameter tuning and feature selection, the classifier obtained accuracy, sensitivity, and specificity ranging between 90 to 92 percent. Out of the total 298 features, 20 features were selected as the most significant factors that affect SWB in young Chinese medical students. These features were related to perception towards life, reaction to failure, sleep quality, and interpersonal relationships. Based on these findings, the authors recommend a brief questionnaire comprising 20 questions to analyze the state of SWB in adolescents with considerable efficiency. While the study makes a valuable contribution towards management of psychological well-being of growing adolescents, the results may not be generalizable on other ethnicities or students studying in different academic programs.

2.1.2 Behavioral Problems

Tate et al. utilized the Strengths and Difficulties Questionnaire (SDQ) for prediction of the risk of development of behavioral problems in adolescence [33]. The researchers designed a time-series analysis where data is examined over the course of a specific time period. The participants included 7,638 Swedish twins from the Child and Adolescent Twin Study (CATSS). This provided clinical, socioeconomic, and environmental data of the participants collected previously at ages 9 and 12. The SDQ was administered to parents of participants who were presently 15 years old. Five machine learning algorithms viz. Random Forest (RF), XGBoost, Logistic Regression (LR), Artificial Neural Network (ANN) and Support Vector Machine (SVM) were used to develop predictive models. Performance of each model was evaluated using area under the receiver operating characteristic curve (AUC), which revealed that RF, SVM, and LR are the most accurate with AUCs above 0.7. Due to mediocre accuracies and specific demographics of the study, the efficiency and generalizability of the developed models necessitates further research and improvement for practical implementation.

2.1.3 Developmental Issues due to Political Conflicts

Qasrawi et al. used results of screening and diagnostic tests, along with socioeconomic and environmental information of 6,373 Palestinian children and adolescents aged 10 to 15 years [34]. Their main objective was to outline risk factors that negatively impact the social and cognitive development of children experiencing political violence and turmoil. Prediction of impaired cognition and well-being was a secondary aim. GB, SVM, RF, ANN, k-Nearest Neighbor (kNN), and Decision Tree (DT) were used to develop models to assess these risk factors, where RF exhibited the best accuracy of 0.91. The Gini importance of the features calculated by the top-performing RF model were used to elucidate the significant risk factors, which included maltreatment, exposure to or participation in violent activities, suicide attempts, tobacco abuse, PTSD, anxiety, and depression. Through these findings, the authors outlined areas of concern which should be emphasized in management and treatment strategies for the preservation of cognitive development of growing children exposed to political violence at a young age. The accuracy of the top-performing model is also satisfactory. However, since the main objective of the study was not the prediction of cognitive issues, it is not appropriate to scrutinize the predictive efficacy and performance of the model in a clinical scenario.

2.2 ML for Specific Disorders

2.2.1 Attention Problems, Anxiety, Academic Problems, ADHD, and PDD

Sumathi et al. acquired the clinical data of 60 children and adolescents from a psychologist in India [35]. Eight algorithms viz. AODEsr, Functional Tree (FT), RBF Network, IB1 Classifier, Kstar Classifier, Multiclass Classifier (MCC), Multilayer Perceptron (MLP), and LADTree were used to develop predictive models for Attention Problems, Anxiety Problems, Academic Problems, Attention-Deficit/Hyperactivity Disorder (ADHD), and Pervasive Developmental Disorder (PDD). Among these, MCC, MLP, and LADTree had the best performance, with accuracies ranging between 0.8 and 0.9. While the performance of the models is optimal for clinical use, the data used to train them is insufficient. Moreover, only data of diagnosed patients is used. Typically-developing children and adolescents with no mental health diagnoses have not been incorporated. This discrepancy might have caused over-fitting of the models for a clinical diagnosis. The sensitivity and specificity of such a model on real-world data consisting of individuals with normal psychosocial development might be unsatisfactory.

2.2.2 Generalized and Separation Anxiety Disorders

Carpenter et al. utilized the Preschool Age Psychiatric Assessment (PAPA) to develop an ML-based screening tool for assessing the risk of generalized anxiety disorder (GAD) and separation anxiety disorder (SAD) in pre-school children [36]. The PAPA was administered to 1,224 children aged 2 to 5 years. This data was used to train and test alternating decision trees (ADTrees). For SAD and GAD, optimal metrics were achieved using 17 and 7 of the original 34 items of the PAPA, respectively. The accuracies for the detection of both disorders were above 96%. Based on these results, the number of questions in PAPA could be reduced to the most significant predictors of children at risk of developing GAD or SAD, potentially resulting in a less time-consuming and complex screening process. While the metrics of the predictive models are quite remarkable, the models warrant validation on other ethnic cohorts to improve their generalizability.

2.2.3 Post-Traumatic Stress Disorder

Ge et al. identified factors that contribute in the development of Post-Traumatic Stress Disorder (PTSD) in children in earthquake-stricken areas [37]. Clinical, socioeconomic, and environmental information was analyzed using XGBoost. After feature selection and optimization of parameters, accuracy of the model was 74.5%. Age, gender, nature of lifestyle, environmental quality of neighborhood and residence, and sleep quality were recognized as the most significant features. From a commercial and practical aspect, the deployment of this model may be quite challenging. Firstly, the mental health issue being detected is quite specific, as the study population is limited to children affected by earthquakes. While these results may be reproducible upon children who have suffered through other natural disasters like floods, tsunamis, and cyclones, the model may not be fit to detect PTSD in other cohorts. Additionally, the metrics of the developed model are not efficient enough to be adopted in a clinical practice. However, the significant factors identified in the study could be used to make informed decisions for the rehabilitation of children exposed to natural disasters.

2.2.4 Depression and Anxiety

Qasrawi et al. used RF, ANN, DT, SVM, and Naïve Bayes to develop predictive models for the detection of depression and anxiety in Palestinian children and young adolescents [38]. Clinical, socioeconomic, and environmental data of 3,984 schoolchildren was used to train and test these algorithms. SVM produced the best predictive models with accuracies above 92%. Most significant features have also been explained which included exposure to violence, bullying, socioeconomic status, and academic performance. The metrics of the proposed model are quite impressive. However, the demographics of the study render the reproducibility and generalizability of these results debatable. Not only do the children belong to a specific ethnic group, they also reside in an occupied territory burdened with political and social turmoil. The nuances of the well-being of children exposed to such dire circumstances are not comparable with children living in peaceful and independent states.

2.3 Gaps in the Literature

The strengths and limitations of the studies constituting the literature review are described in Table 2.1.

Table 2.1: Summary of the literature review.

No.	Authors	Mental Health Issue	Strengths	Limitations
1	Zhang et al. (2019)	Subjective Well-Being (SWB) and associated factors	- Model performance metrics between 90-92% - Large dataset (10,518 instances)	- Lack of generalizability (Chinese cohort) - Older age group (Freshmen and sophomore medical students)
2	Tate et al. (2020)	Behavioral Problems	- Model AUC above 0.7 - Large dataset (7,638 instances)	- Lack of generalizability (Swedish cohort) - Older age group (9 to 15 years)

3	Qasrawi et al. (2023)	Developmental issues due to Political Conflicts	- Model accuracy = 0.91 - Large dataset (6,373 instancess)	- Lack of generalizability (Palestinian cohort) - Older age group (10 to 15 years)
4	Sumathi et al. (2016)	Attention Problems, Anxiety, Academic Problems, ADHD, PDD	- Model accuracies between 0.8 and 0.9 - Vast age range (3 to 15 years)	- Lack of generalizability (Indian cohort) - Limited sample size (60 instances)
5	Carpenter et al. (2016)	Separation Anxiety Disorder and Generalized Anxiety Disorder	- Model accuracies above 96% - Large dataset (1,224 instances)	- Lack of generalizability (American cohort) - Younger age group (2 to 5 years)
6	Ge et al. (2019)	Post-Traumatic Stress Disorder (PTSD)	- Model accuracy = 74% - Vast age range (5 to 17 years) - Large dataset (2,099 instances)	- Lack of generalizability (Chinese cohort)
7	Qasrawi et al. (2022)	Depression and Anxiety	- Model Accuracies above 92% - Large dataset (3,984 instances)	- Lack of generalizability (Palestinian cohort) - Older age group (10 to 15 years)

At the time of surveying literature, no published work focusing on the prediction of mental health issues of Pakistani children and adolescents was observed. This is recognized as a significant research gap, as efficient smart screening processes optimized for Pakistani youth are needful to overcome the pervasive challenges hampering mental health intervention strategies. The influence of sociodemographic factors on the development of mental health issues discusses in Section 1.1.1 further evidences the significance of this research gap.

Most of the studies focused on participants in their late childhood and adolescence (ages 10 and above). Younger children (ages 9 and below) are not as prevalent in the study populations. The emphasis on older age groups may have occurred by default due to easier availability of data of older children and adolescents, or it may be deliberate as adolescence is a critical phase for monitoring psychological problems [1]. However, early-onset psychopathological conditions do exist which begin persisting from ages as young as 2 years [39]. Therefore, another research gap is highlighted where performance of AI-driven predictive models for mental health issues of younger children should be investigated as well.

The utilization of local data of Pakistani children and adolescents aged 6 to 17 years in the present study will address these research gaps and provide significant contributions within interdisciplinary research between machine learning and mental health.

3. METHODOLOGY

This study followed a conventional ML pipeline of data collection, pre-processing, feature selection, and training of ML algorithms. A secondary local dataset is procured by an online mental health clinic, which is pre-processed and labelled according to the scoring guidelines of the PSC. This is followed by a comprehensive feature selection process involving statistical and computational approaches. Lastly, the data are used to train and test six ML algorithms that encompass classical, ensemble, deep learning approaches. The performance metrics of each algorithm are evaluated to deduce the best performing model that can be transformed into a decision support system. The overall workflow is summarized in Figure 3.1.

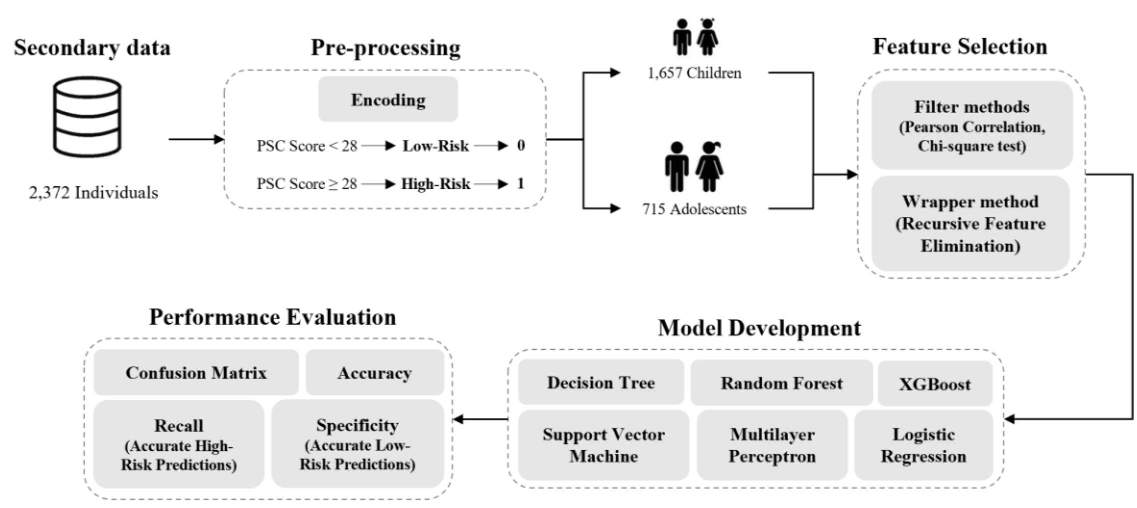

Figure 3.1: Overall workflow of the research methodology.

3.1 Data Collection

Data of 2,372 children and adolescents is provided by the online mental health clinic SehatYab based in Islamabad, Pakistan [40]. An online survey was promoted through social media and data was collected using the Urdu version of the parent-rated PSC over the course of 6 months (31st March, 2022 to 30th September, 2022). Parents of children and adolescents aged 6 to 17 years participated in the survey which consisted of the 35 items of the PSC, where each item corresponds to 3 options; Never, Sometimes, and Often. These options translate to the numeric scores of 0, 1, and 2, respectively. The scores of all individual items are added to deduce the final result of the PSC. A score below the recommended cut-off point of 28 indicates that the risk of psychosocial dysfunction is low. If the score is equal to or above 28, the risk of psychosocial dysfunction is high and professional intervention is recommended.

The 35 items of the PSC are divided into 3 subscales, where the attention and internalizing subscales comprise 5 questions each, and the externalizing subscale consists of 7 questions. Cut-off scores for these subscales are also defined. For the attention subscale and externalizing subscales, if the score is equal to or greater than 7, the risk for attention and externalizing problems is high. For the internalizing subscale, a score of 5 or above indicates high risk of internalizing problems. In addition to the 35 items of the PSC, age and city of residence of the participants were also recorded.

3.2 Data Pre-processing

All pre-processing is done using the software SPSS (Version 20) and programming language Python (Version 3.10) [41], [42]. Missing value analysis revealed no missing values in the data. According to the recommendations of psychiatrists at SehatYab and the cut-off scores explained earlier, subjects with a total score of 27 or below on the PSC are labelled “Low-Risk” and those with a total score of 28 or above are labelled “High-Risk” for psychosocial dysfunction.

Similarly, scores of the 3 subscales of the PSC are analyzed to recognize subjects at risk of attention, internalizing, and/or externalizing problems. For the attention and externalizing subscales, subjects with a score of 6 or below are labelled “Low-Risk” and those with a score of 7 or above are labelled “High-Risk” for attention and externalizing problems. For the internalizing subscale, instances with a score of 4 or below are labelled “Low-Risk” and those with a score of 5 or above are labelled “High-Risk” for internalizing problems. All “Low-Risk” and “High-Risk” labels are numerically recoded to 0 and 1, respectively. This is done to generate numeric categories or targets in the data for subsequent application of ML algorithms. The resulting data consists of 45 features (Table 3.1).

Data are divided into 2 age groups based on the age stages defined by World Health Organization (WHO) [43]. These age stages are listed in Table 3.2. The resulting groups constitute 1,657 children aged 6 to 11 years, and 715 adolescents aged 12 to 17 years. All subsequent steps are performed separately on both groups of data to analyze any potential differences between children and adolescents. For subsequent feature selection and machine learning, age of the subjects and the 35 items of the PSC are chosen as predictive variables for the risk of psychosocial dysfunction only.

Table 3.1: Data description after pre-processing.

No.	Feature Name	Description	Type	Values
1	Age	Age of the subject	Quantitative	Ranging from 6 to 17 years
2	City	City of Residence	Qualitative	Names of 89 cities
3 - 37	PSC1 – PSC35	Items of the PSC	Qualitative	"Never" = 0, "Sometimes" = 1, "Often" = 2
38	Total_Score	Total Score of all 35 items	Quantitative	Ranging from 0 to 64
39	Total_Score_ Binary	Scores ≤ 27 labelled "Low-Risk" and scores ≥ 28 labelled "High-Risk"	Qualitative	"Low-Risk" = 0, "High-Risk" = 1
40	Att_Score	Score of the 5 items of the Attention Subscale	Quantitative	Ranging from 0 to 10
41	Att_Score_ Binary	Scores ≤ 6 labelled "Low-Risk" and scores ≥ 7 labelled "High-Risk"	Qualitative	"Low-Risk" = 0, "High-Risk" = 1
42	Int_Score	Score of the 5 items of the Internalizing Subscale	Quantitative	Ranging from 0 to 10
43	Int_Score_ Binary	Scores ≤ 4 labelled "Low-Risk" and scores ≥ 5 labelled "High-Risk"	Qualitative	"Low-Risk" = 0, "High-Risk" = 1
44	Ext_Score	Score of the 7 items of the Externalizing Subscale	Quantitative	Ranging from 0 to 14
45	Ext_Score_ Binary	Scores ≤ 6 labelled "Low-Risk" and scores ≥ 7 labelled "High-Risk"	Qualitative	"Low-Risk" = 0, "High-Risk" = 1

Table 3.2: Age stages recommended by World Health Organization.

Age Stage Descriptor	Age Groups
Toddler	1 to < 2 year
Early Childhood	2 to < 6 years
Middle Childhood	6 to < 11 years
Early Adolescence	11 to < 16 years
Late Adolescence	16 to <21 years

3.3 Data Analysis

3.3.1 Descriptive Analysis

To describe characteristics of the data, demographics including age and city of residence of the study population are analyzed. Prevalence of psychosocial dysfunction, attention problems, internalizing problems, and externalizing problems according to the PSC are also determined using the scoring instructions of Massachusetts General Hospital. Measures of central tendency including mean and mode are calculated for age of the subjects, total score of the PSC, and scores of the three subscales of the PSC for each age group.

3.3.2 Internal Consistency and Reliability Analysis

To the best of our knowledge, only the self-rated version of the PSC has been administered and validated on Pakistani pediatric populations [44]. As this study aims to utilize the parent-rated version of the PSC, its internal consistency and reliability as a suitable questionnaire for Pakistani children and adolescents must be analyzed. For this, reliability analysis is performed in SPSS by calculating Cronbach’s alpha and evaluating inter-item correlations.

Cronbach's alpha is a statistical assessment measure for internal consistency, reliability, and stability of questionnaires, particularly those pertaining to social, psychological, and biological sciences [45]. It indicates how well the items of a questionnaire relate to one another. This relationship is determined by calculating inter-item correlations among all items and the magnitude of Cronbach’s alpha. Typically, the magnitudes of these correlations should be within the range of 0.2 to 0.6. Values less than 0.2 indicate very weak correlation among items, meaning that the questions are dissimilar and are diverging from the main focus of the questionnaire. Values greater than 0.6 indicate a strong correlation which suggests that the items are very similar to each other. This renders them repetitive and redundant. Correlations between 0.2 and 0.6 ensure that the items of the questionnaire are broad enough to sufficiently encompass the scope of the questionnaire while avoiding repetition. In SPSS, the inter-item correlations will be computed as a 35x35 matrix, which will be difficult to interpret. Therefore, the inter-item correlations have been computed as heatmaps in Python using the same statistical methods to ensure better and easier visualization. The magnitude of Cronbach's alpha ranges from 0 to 1, where higher values indicate greater internal consistency, suggesting that the items are measuring the same underlying concept or domain of the questionnaire. Generally, a value between 0.7 and 0.9 is considered acceptable [46].

3.3.3 Hypothesis Testing

Hypothesis testing is a fundamental statistical concept of inferential analysis. It is used to estimate population parameters based on sample data and explain the significance of an effect or relationship [47]. Hypothesis testing begins with formulation of two competing hypotheses. One hypothesis deals with the scenario that no effect or relationship exists or can be inferred. This is the null hypothesis (H_0) which is to be disproved. The second hypothesis is the alternative hypothesis (H_1 or H_a) which states that the relationship or effect being investigated exists. Determination of the correct hypothesis involves estimation of statistical significance. For this, statistical test and a level of significance is chosen. The choice of the test is determined by the characteristics of the data and the nature of inferential analysis. The level of significance (α) is generally set at 0.05. This indicates that there is a 5% chance of committing an error in selection of the correct hypothesis. Statistical tests compute a test statistic with a corresponding p-value which reflects the statistical significance of the computation. If this p-value is less than α, the null hypothesis is rejected. Therefore, the relationship being investigated exists. If p-value is greater than α, null hypothesis cannot be rejected, and therefore the claim of the researcher is disproven.

In this study, potential differences between the results of children and adolescents are investigated via hypothesis testing in SPSS. The following null and alternative hypotheses are formulated:

***H_0*:** There is no difference between the results of children and adolescents.

***H_1*:** There is a difference between the results of children and adolescents.

Independent samples t-test is selected as the test statistic as the two groups being compared (children and adolescents) are independent of each other. The significance level is set to 0.05 and the mean scores of children and adolescents are compared. Age is selected as the grouping variable in this case, where subjects below the age of 12 constitute the "Children" age group and 12-year-old or older subjects make up the "Adolescent" age group.

3.4 Feature Selection

Prior to the implementation of ML algorithms, two methods of feature selection are applied to identify the most significant of the 35 independent variables. These include two filter methods viz. Chi-square test and Pearson Correlation, and one wrapper method i.e. Recursive Feature Elimination (RFE).

3.4.1 Filter Methods

Filter methods calculate a statistic based on the intrinsic properties of features to explain their significance [48]. Pearson Correlation evaluates a linear relationship between continuous quantitative variables by calculating the Pearson Correlation Coefficient R [49]. Its value ranges between -1 to +1. Negative values indicate an inverse relationship meaning that increase in one variable decreases the other, and positive values indicate a direct relationship where increase in one variable also increases the other. The magnitude of R indicates strength of the relationship. Irrespective of the sign, values closer to 1 indicate a strong relationship and values near 0 show weak correlation. Typically, the criteria explained in Table 3.3 is followed while deciding the strength of a relationship according to the magnitude of R [50]. In addition to R, a p-value is also calculated which establishes the statistical significance of the correlation. A p-value below the selected level of significance suggests that the results are not significant and p-values above it indicate significant results.

Table 3.3: Criteria for determining the strength of a relationship according to the magnitude of Pearson Correlation Coefficient.

Absolute magnitude of R	Interpretation
0.00 – 0.10	Negligible/Very weak relationship
>0.10 – 0.39	Weak relationship
0.40 – 0.69	Moderate relationship
0.70 – 0.89	Strong relationship
0.9 – 1	Very strong relationship

For this research, SPSS is used to compute R for the variable "age" as it is a quantitative variable. The level of significance is set to 0.05. Variables with insignificant results and/or weak relationships are investigated during subsequent model development.

Chi-square tests are a statistical method that can be used to analyze the association between qualitative or categorical variables [51]. Similar to Pearson Correlation, Chi-square tests also involve the calculation of a statistic (χ^2) and a p-value for significance. In order to determine the strength of the association, additional statistics called Contingency Coefficients (C) can be computed. Generally, values equal to below 0.1 are considered to represent weak associations [52]. In this study, Chi-square tests are applied and C are computed for the remaining 35 categorical variables (items of the PSC) with a significance level of 0.05. Variables with magnitudes of C below 0.1 and/or p-values below 0.05 are investigated during subsequent model development.

3.4.2 Wrapper Methods

Wrapper methods incorporate an ML algorithm to determine which subset of features produces the best performance metrics [53]. The selection criterion is based on the feature importance calculated by the ML algorithm during training. In the end, the subset of features resulting in the highest classification accuracy is selected. Recursive Feature Elimination (RFE) is an example of a wrapper method where subsets of features are reiteratively or *recursively* tested and the combination of features in changed in each iteration [54]. Using Python, RFE is applied in this study to reiteratively test different subsets and combinations of the 36 features and select those features which produce the best performance metrics. Random Forest (RF-RFE) and XGBoost (XGB-RFE) are applied as estimators. Tree-based algorithms are selected to ensure that similar evaluation criterion (Gini importance) is used for feature selection.

3.5 Model Development

The 35 items of the PSC and age of the subjects are chosen as the independent variables and the categorical result of “Low-Risk” and “High-Risk” subjects for psychosocial dysfunction is selected as the target variable. Data are split into 80% training data and 20% test data. Six ML algorithms are applied to build predictive models using the PSC. Primarily, default parameters as defined by the machine learning Python library ***scikit-learn*** are used [55]. Some parameters are altered to optimize the performance of the models.

3.5.1 Decision Tree

Decision trees are one of the most classical ML algorithms for supervised learning. A decision tree (DT) constructs a tree-like model or flowchart for classification or regression. Each feature or variable corresponds to a model. It begins with a root node which splits or branches out into leaf nodes on the basis of the Gini impurity calculated for each variable. Features of the data are recursively split until an end-point is reached where all data points are categorized into the specified target classes [56]. DT in this study utilize the *best* splitter and *gini* criterion for splitting each node into appropriate leaf nodes. The default maximum allowed depth is *none* so that growth of the tree is unrestricted. The minimum sample split is set at 10 samples.

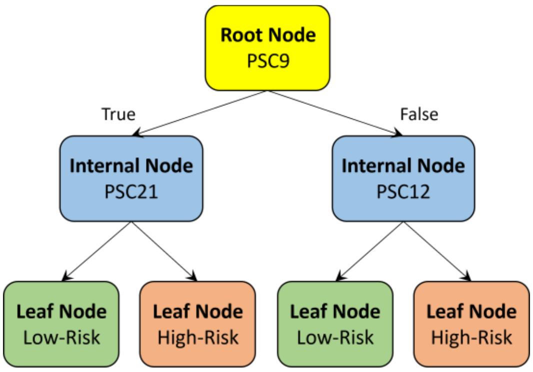

Figure 3.2: A simplified example of Decision Trees in this study.

3.5.2 *Random Forest*

Random Forest (RF) is an ensemble technique which uses multiple DTs for classification or regression. They are an extension of decision trees which is more accurate, less prone to over-fitting, and capable of handling larger amounts of data. Different subsets of the data are generated via bagging and each subset is used to train different decision trees. Splitting of nodes depends on the node size, number of estimators, and number of features. The final output is estimated by majority voting or calculating the average performance of all trees [57]. In this study, RFs of 100 DTs are constructed with *gini* criterion, minimum sample split set to 10, unrestricted maximum depth, and *auto* for consideration of maximum features at each split.

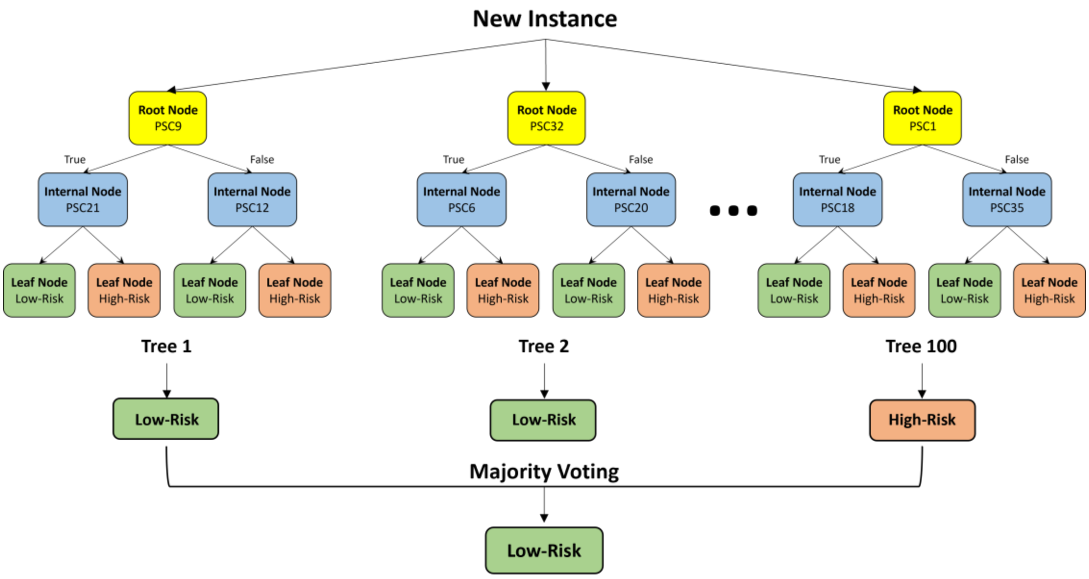

Figure 3.3: A simplified representation of Random Forests in this study.

3.5.3 *XGBoost*

Also known as Extreme Gradient Boosting, XGBoost (XGB) applies a sequential, multi-threaded approach to gradient boosting to train multiple decision trees in parallel. During training, errors of the previous tree are rectified by the next one. Weighted average of the performance of all trees is used to estimate the final output [58]. XGB is able to handle large data sets and prevent over-fitting via regularization. In this study, XGBs of 100 trees, maximum depth with a default value of 3, default learning rate of 0.1, *gbtree* booster, and *binary:logistic* objective are developed.

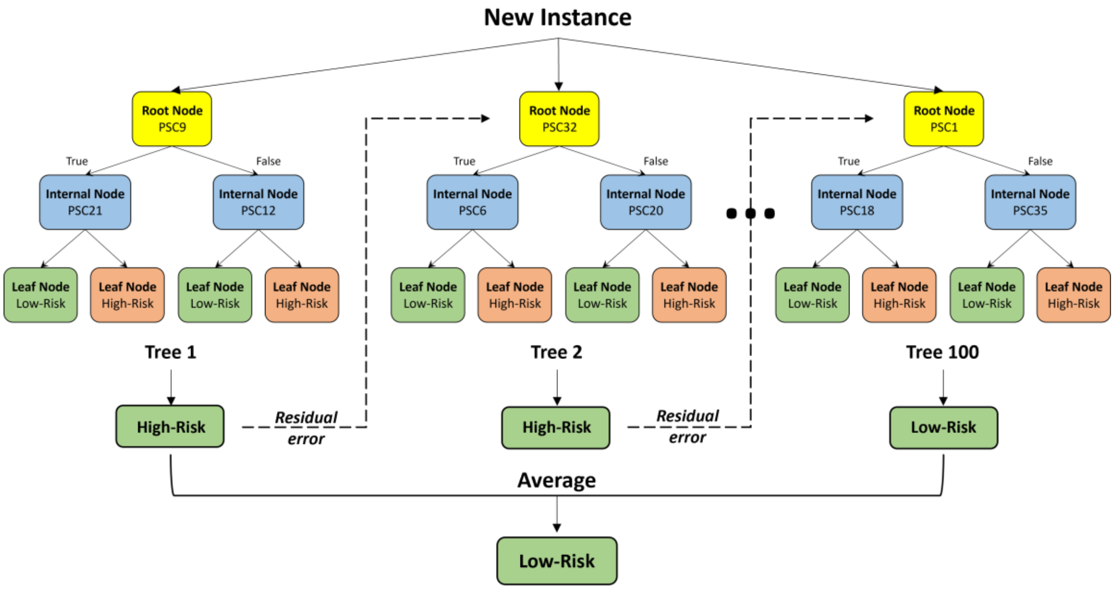

Figure 3.4: A simplified representation of XGBoost models in this study.

3.5.4 Support Vector Machine

A support vector machine (SVM) generates hyperplanes and support vectors for classification or regression. A hyperplane is a high-dimensional line that separates data points into categories. These data points are referred to as support vectors. The hyperplane with the widest margin or distance from the support vectors is selected as the optimal decision boundary [56]. Data are classified based on which side of the decision boundary it falls on. For this study, *Linear* kernel has been chosen and probability estimates are enabled and calculated using *CalibratedClassifierCV*. The remaining default parameters include a trade-off oof 1.0 between margin maximization and allowed miscalculations and tolerance of 0.001.

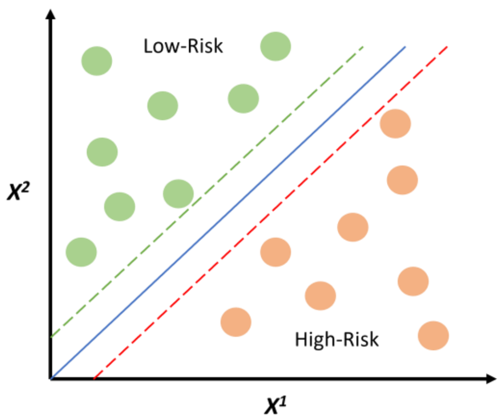

Figure 3.5: Graphical representation of the Support Vector Machine.

3.5.5 Multilayer Perceptron

Multilayer Perceptron (MLP) is an Artificial Neural Network which consists of interconnected neurons arranged in multiple layers [59]. These layers constitute:

- An input layer, where data points are fed or introduced into the neural network.
- A single or multiple hidden layer(s), which perform a series of calculations using non-linear activation functions to recognize and learn patterns in the training data. These learned patterns are then used for classification or prediction on test data.
- An output layer, where appropriate target variables are designated to the input.

Each neuron is assigned a weight during the training process, which is optimized according to the accuracy and significance of the neuron in classification or prediction. The architecture of MLPs in this study comprises a feed-forward neural network of 2 hidden layers with 64 neurons in the first layer and 32 neurons in the second layer. The number of iterations has been set to 1000 with *relu* activation function.

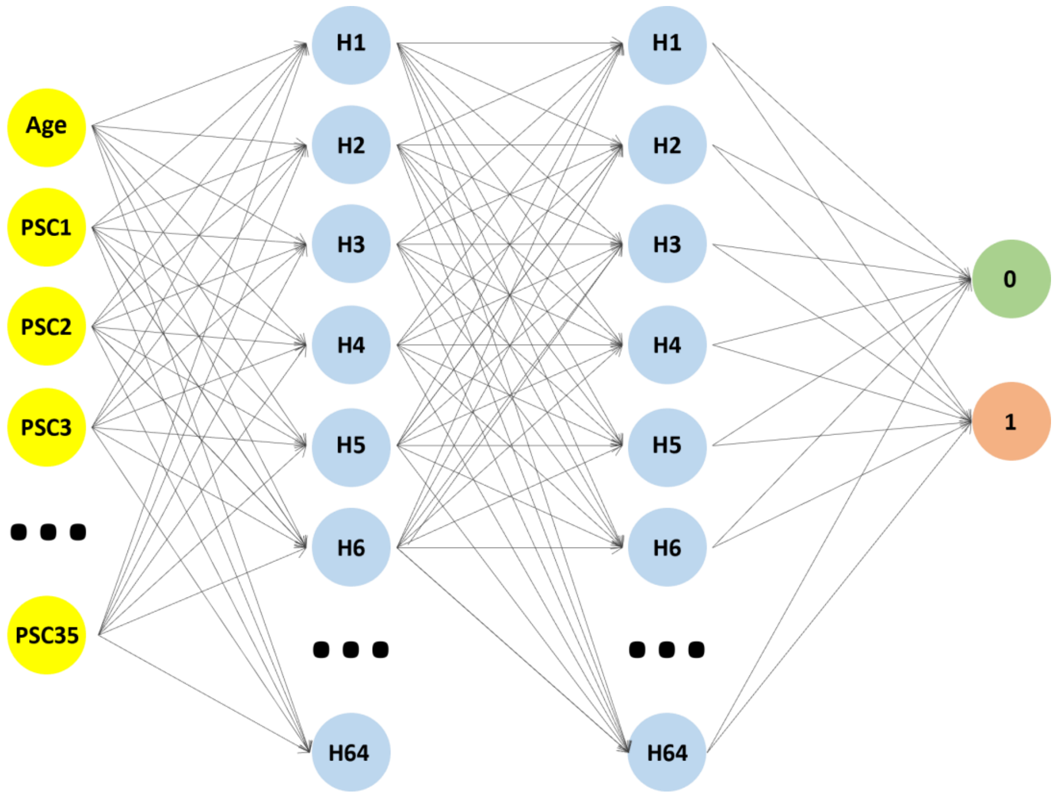

Figure 3.6: Visualization of the Multilayer Perceptron.

3.5.6 Logistic Regression

Logistic Regression (LR) is frequently used in the fields of medicine and social sciences [60]. It uses the logit function to estimate the probability of an outcome based on one or multiple independent factors [56]. Traditionally, LR performs binary tasks where the events are designated as ‘0’ and ‘1’. In this study, a model with 100 iterations, *L2* penalty term, a default regularization strength of 1.0, a tolerance of 0.001, and *lbfgs* solver has been developed.

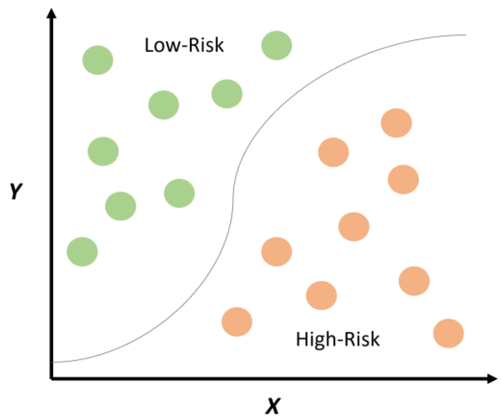

Figure 3.7: Graphical representation of Logistic Regression.

3.6 Performance Evaluation

In order to ascertain the efficacy of the proposed models, various aspects or metrics of their performance are evaluated. These metrics are evaluated during the training and testing phase, where classification labels determined by the model are compared with the actual targets labelled in the dataset. For this study, the performance metrics of the models are evaluated at two stages:

i. Evaluation of the most significant set of features by training and testing the algorithms on the original dataset and the subsets of features provided after feature selection.
ii. Determination of the most accurate predictive model by comparing the performance of each ML algorithm.

The following metrics are chosen to provide a comprehensive account of the performance of the ML algorithms:

3.6.1 Confusion Matrix

A confusion matrix compares the predicted classes of the model against actual classes of the data. Rather than being considered a performance metric, it is more a tabular representation of the predictive performance of the algorithm. The rows of a confusion matrix represent the actual classes of the data and the columns correspond to the predicted classes labelled by the algorithm. In this study, the class labels are "Low-Risk" and "High-Risk" which correspond to "Negative" and "Positive", respectively. The resulting confusion matrix will assume the form defined in Table 3.4.

Table 3.4: Confusion matrix for the binary classification problem in the study.

Class labels		Predicted	
		Low-Risk	High-Risk
Actual	Low-Risk	True Negative	False Positive
	High-Risk	False Negative	True Positive
Negative = Low-Risk, Positive = High-Risk			

3.6.2 Accuracy

Accuracy is the measure of the number of accurate predictions out of total predictions. It is the most basic performance metric and is calculated by dividing the number of correct predictions by all predictions.

$$Accuracy = \frac{Number\ of\ Correct\ Predictons}{Number\ of\ Total\ Predictions} \tag{3.1}$$

3.6.3 Precision

It is a measure of the number of accurate positive predictions out of all positive predictions of the model. In this study, precision will define the number of High-Risk predictions accurately identified by the model by comparing the number of actual High-Risk cases with the total number of High-Risk instances predicted by the model.

$$Precision = \frac{True\ Positives\ (Actual\ High - Risk)}{True\ Positives + False\ Positives\ (Predicted\ High - Risk)} \tag{3.2}$$

3.6.4 Recall

Recall measures the number of accurate positive predictions of the model. This metric is also referred to as "sensitivity" in health sciences. In this study, recall will be the ratio between true High-Risk predictions and actual High-Risk cases in the data.

$$Recall = \frac{True\ Positives\ (Actual\ High - Risk)}{True\ Positives + False\ Negatives\ (Total\ High - Risk)} \quad (3.3)$$

3.6.5 Specificity

Specificity measures the number of accurate negative predictions made by the model. In this study, specificity will be calculated by comparing the true Low-Risk predictions of the model with all Low-Risk cases present in the data.

$$Specificity = \frac{True\ Negative\ (Actual\ Low - Risk)}{True\ Negatives + False\ Positives\ (Total\ Low - Risk)} \quad (3.4)$$

4. RESULTS AND DISCUSSIONS

This chapter describes the results obtained from the proposed methodology and discusses them in light of similar literature. The strengths, limitations, and future recommendations from the outcomes of this study are also highlighted.

4.1 Descriptive Analysis

4.1.1 Demographics

The ages of children ranged from 6 to 11 years, where the mean age was 8.14 ± 1.66 years. 6 years was the most abundant age (N=354) and 11 was the least abundant age (N=186). The ages of adolescents ranged from 12 to 17 years with a mean age of 14.04 ± 1.82 years. 12 years was the most abundant age (N=202) and 15 was the least abundant age (N=67). Detailed age distribution of the data is provided in Table 4.1.

Table 4.1: Age Distribution of the data of children and adolescents.

	Children (N=1,657)			Adolescents (N=715)		
	Year	**N**	**%**	**Year**	**N**	**%**
Age	6	354	21.4	12	202	28.3
	7	327	19.7	13	132	18.5
	8	316	19.1	14	125	17.5
	9	240	14.5	15	67	9.4
	10	234	14.1	16	70	9.8
	11	186	11.2	17	119	16.6
Mean of Age	8.4			14		
Mode of Age	6			12		

Data were collected from 89 cities of Pakistan. Majority of the participants are from big cities like Karachi (N=592), Islamabad (N=391), Lahore (N=373), Multan (N=103), and Faisalabad (N=98). This may be attributed to better facilities and outreach in these areas. Number of participants from all cities are listed in Table 4.2.

Table 4.2: City-wise Data Distribution.

No.	City	Frequency	No.	City	Frequency	No.	City	Frequency
1	Karachi	592	31	Kasur	9	61	Kot Addu	3
2	Islamabad	391	32	Saddiqabad	9	62	Mansehra	3
3	Lahore	373	33	Toba Tek Singh	9	63	Charsadda	2
4	Multan	103	34	Bhakkar	8	64	Chenab Nagar	2
5	Faisalabad	98	35	Gilgit	8	65	Kalat	2
6	Rawalpindi	68	36	Jhang City	7	66	Kamalia	2
7	Gujranwala	64	37	Mirpur Khas	7	67	Kandhkot	2
8	Hyderabad City	51	38	Nawabshah	7	68	Karak	2
9	Peshawar	40	39	Bannu	6	69	Khushab	2
10	Sargodha	39	40	Chiniot	6	70	Lala Musa	2
11	Bahawalpur	33	41	Hassan Abdal	6	71	Mingaora	2
12	Kulachi	30	42	Muzaffarabad	6	72	Pasrur	2
13	Sialkot City	30	43	Nowshera	6	73	Samundri	2
14	Gujrat	27	44	Saidu Sharif	6	74	Turbat	2
15	Quetta	23	45	Bahawalnagar	5	75	Umarkot	2
16	Sahiwal	22	46	Bhalwal	5	76	Chaman	1
17	Rahimyar Khan	19	47	Gojra	5	77	Gwadar	1
18	Abbottabad	18	48	Larkana	5	78	Hangu	1
19	Attock Khurd	15	49	Nankana Sahib	5	79	Jacobabad	1
20	Mandi Bahauddin	15	50	New Mirpur	5	80	Kahror Pakka	1
21	Mianwali	15	51	Ahmadpur East	4	81	Kotli	1
22	Kabirwala	11	52	Khanpur	4	82	Mian Channun	1
23	Mardan	11	53	Kundian	4	83	Pattoki	1
24	Okara	11	54	Murree	4	84	Risalpur	1
25	Pakpattan	11	55	Muzaffargarh	4	85	Sambrial	1
26	Vihari	11	56	Shakargarh	4	86	Swabi	1
27	Dera Ismail Khan	10	57	Chishtian	3	87	Tando Allahyar	1
28	Shekhupura	10	58	Harunabad	3	88	Tank	1
29	Sukkur	10	59	Kharian	3	89	Zhob	1
30	Dera Ghazi Khan	9	60	Kohat	3			

4.1.2 Prevalence of Psychosocial Problems according to the PSC

Table 4.3 describes the Low-Risk and High-Risk individuals identified for psychosocial dysfunction, attention issues, internalizing problems, and externalizing problems according to the scoring system of the PSC.

Table 4.3: Prevalence of psychosocial problems in the study populations according to the Pediatric Symptom Checklist.

Psychosocial Problems	Children (n=1,657)		Adolescents (n=715)	
	Low-Risk	High-Risk	Low-Risk	High-Risk
	n (%)	n (%)	n (%)	n (%)
Psychosocial Dysfunction	855 (51.6)	802 (48.4)	260 (36.4)	455 (63.6)
Attention Problems	1,273 (76.8)	384 (23.3)	471 (65.9)	244 (34.1)
Internalizing Problems	1,174 (70.9)	483 (29.1)	368 (51.5)	347 (48.5)
Externalizing Problems	876 (52.9)	781 (47.1)	358 (50.1)	357 (49.9)

4.1.2.1 Psychosocial Dysfunction

51.6% (N=855) of children scored 27 or below on the PSC, indicating that they are at low risk for psychosocial dysfunction. The remaining 48.4% (N=802) are at high risk as they scored 28 or above on the PSC. Based on these proportions, it can be said that psychosocial dysfunction is somewhat prevalent among children in the study population. Among adolescents, 36.4% (N=260) scored 27 or below on the PSC, indicating that they might not be psychologically impaired. The remaining 63.6% (N=455) might be psychologically impaired as they scored 28 or above on the PSC. This indicates that psychosocial dysfunction is significantly prevalent among adolescents in the study population.

4.1.2.2 Attention Problems

Among children, 76.8% (N=1,273) scored 6 or below on the Attention Subscale of the PSC, indicating that they might not have attention issues. The remaining 23.2% (N=384) might have attention issues as they scored 7 or above on the Attention Subscale of the PSC. 65.9% (N=471) of the adolescents scored 6 or below on the Attention Subscale of the PSC, indicating that they might not have attention issues. The remaining 34.1% (N=244) might have attention issues as they scored 7 or above on the Attention Subscale of the PSC. Therefore, it is inferred that attention problems are not prevalent in the overall study population.

4.1.2.3 Internalizing Problems

Among children, 70.9% (1,174) scored 4 or below on the Internalizing Subscale of the PSC, indicating that they might not have internalizing issues. The remaining 29.1% (483) might have internalizing issues as they scored 5 or above on the Internalizing Subscale of the PSC. Therefore, internalizing problems are also not prevalent among children in the study population. Among adolescents, 51.5% (368) scored 4 or below on the Internalizing Subscale of the PSC, indicating that they might not have internalizing issues. The remaining 48.5% (347) might have internalizing issues as they scored 5 or above on the Internalizing Subscale of the PSC. Based on these proportions, it can be said that internalizing issues are somewhat prevalent among the adolescents in the study population.

4.1.2.4 Externalizing Problems

52.9% (876) of the children scored 6 or below on the Externalizing Subscale of the PSC, indicating that they might not have externalizing issues. The remaining 47.1% (781) might have externalizing issues as they scored 7 or above on the Externalizing Subscale of the PSC. Among adolescents, 50.1% (358) scored 6 or below on the Externalizing Subscale of the PSC, indicating that they might not have externalizing issues. The remaining 49.9% (357) might have externalizing issues as they scored 7 or above on the Externalizing Subscale of the PSC. Overall, it can be said that externalizing issues are prevalent among the study population.

4.2 Internal Consistency and Reliability Analysis

Table 4.4 summarizes the reliability analysis of the PSC and its subscales. Cronbach’s alpha of the PSC for children and adolescents is 0.88 and 0.876, respectively. These are satisfactory values as they are above the recommended value of 0.7 for statistically valid and reliable questionnaires. The average inter-item correlations for subscales of the PSC ranged between 0.2 to 0.4 for both age groups, indicating moderate relationships.

Table 4.4: Summary of Reliability Analysis of the PSC and its subscales.

Reliability Statistics	Children	Adolescents
Cronbach's Alpha	0.88	0.876
Mean Inter-item Correlation for Attention Subscale	0.227	0.252
Mean Inter-item Correlation for Internalizing Subscale	0.396	0.402
Mean Inter-item Correlation for Externalizing Subscale	0.305	0.298

Table 4.5 shows the detailed item statistics of the reliability analysis for the PSC administered to children, which encompasses the item-total correlation of each question of the PSC and the effects of deletion of any item on the scale mean, scale variance, and Cronbach's alpha. The item-total correlations range between 0.2 and 0.5, indicating moderate relationships. Deletion of any item results in negligible variation in the scale mean and variance. Cronbach's alpha decreases by a small magnitude upon deletion of any item, indicating that each item of the PSC is significant to maintain its validity.

Table 4.5: Detailed Item Statistics of the PSC administered to children.

Item	Corrected Item-Total Correlation	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Cronbach's Alpha if Item Deleted
PSC1	.261	27.15	126.611	0.879
PSC2	.306	27.35	124.780	0.878
PSC3	.359	27.06	123.685	0.877
PSC4	.330	26.99	123.617	0.878
PSC5	.468	27.47	123.089	0.875
PSC6	.478	27.46	122.274	0.875
PSC7	.257	27.70	126.294	0.879
PSC8	.227	27.36	126.279	0.880
PSC9	.427	27.04	122.609	0.876
PSC10	.305	27.61	125.042	0.878
PSC11	.519	27.24	121.476	0.874
PSC12	.458	26.63	123.815	0.875
PSC13	.527	27.36	121.622	0.874
PSC14	.493	27.15	121.996	0.874
PSC15	.466	27.51	122.287	0.875
PSC16	.418	27.13	123.356	0.876
PSC17	.289	27.73	126.808	0.878
PSC18	.408	27.59	124.399	0.876
PSC19	.519	27.60	122.624	0.874
PSC20	.445	27.46	123.184	0.875
PSC21	.333	27.55	124.800	0.878
PSC22	.486	27.49	122.287	0.875
PSC23	.266	26.95	125.579	0.879

PSC24	.377	27.58	124.412	0.877
PSC25	.292	27.42	125.370	0.878
PSC26	.250	27.14	126.387	0.879
PSC27	.517	27.42	121.011	0.874
PSC28	.403	27.23	122.871	0.876
PSC29	.504	26.88	121.483	0.874
PSC30	.343	27.24	124.042	0.878
PSC31	.388	27.02	123.718	0.877
PSC32	.379	27.23	123.577	0.877
PSC33	.399	27.25	123.698	0.876
PSC34	.384	27.35	123.784	0.877
PSC35	.376	27.10	123.720	0.877

Table 4.6 shows the detailed item statistics of the reliability analysis for the PSC administered to adolescents. Overall, deletion of any item does not produce significant variations in the scale mean or variance. The item-total correlations range between 0.2 and 0.6, indicating moderate relationships. However, PSC7 and PSC23 have very low correlations of 0.106 and 0.069, respectively. Deletion of these item also increases Cronbach's alpha by a small magnitude. This indicates that item 7 (Acts as if driven by a motor) and item 23 of the PSC (Wants to be with you more than before) do not contribute significantly to its validity in case of administration to adolescents. In retrospect, item 23 has negative results in subsequent steps of feature selection and model development as well, which are discussed in detail in Sections 3.4 and 3.5.

Table 4.6: Detailed Item Statistics of the PSC administered to adolescents.

Item	Corrected Item-Total Correlation	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Cronbach's Alpha if Item Deleted
PSC1	0.328	33.04	134.147	0.874
PSC2	0.415	32.76	132.554	0.872
PSC3	0.378	32.71	133.247	0.873
PSC4	0.255	32.82	134.448	0.876
PSC5	0.412	33.16	132.288	0.872
PSC6	0.537	33.04	129.851	0.87
PSC7	0.106	33.55	137.623	0.878
PSC8	0.459	32.98	130.277	0.871
PSC9	0.581	32.65	130.193	0.869
PSC10	0.136	33.2	136.533	0.879
PSC11	0.556	32.69	131.069	0.87
PSC12	0.41	32.35	134.527	0.873
PSC13	0.527	32.93	129.476	0.87
PSC14	0.608	32.67	129.446	0.868

PSC15	0.286	33.09	133.788	0.875
PSC16	0.522	33.02	130.574	0.87
PSC17	0.202	33.42	136.322	0.876
PSC18	0.397	33.16	132.288	0.873
PSC19	0.509	33.25	130.49	0.87
PSC20	0.46	33.2	131.015	0.871
PSC21	0.347	33.2	132.793	0.874
PSC22	0.5	33.16	130.213	0.87
PSC23	0.069	33.07	138.143	0.879
PSC24	0.513	33.18	130.485	0.87
PSC25	0.202	33.31	135.995	0.876
PSC26	0.28	33.35	134.415	0.875
PSC27	0.506	33.18	129.707	0.87
PSC28	0.436	32.98	130.944	0.872
PSC29	0.463	32.89	131.247	0.871
PSC30	0.359	32.93	132.995	0.873
PSC31	0.34	32.93	133.106	0.874
PSC32	0.376	33.15	132.053	0.873
PSC33	0.46	33.02	130.759	0.871
PSC34	0.196	33.44	136.102	0.877
PSC35	0.385	33.15	132.349	0.873

Overall, the inter-item correlations among all 35 items of the PSC indicate a weak to moderate relationship, where majority of the correlations lie between 0.1 to 0.6 (Figure 4.1). This shows that items of the questionnaire focus on a specific domain (psychosocial dysfunction), but are diverse enough to not be repetitive or redundant.

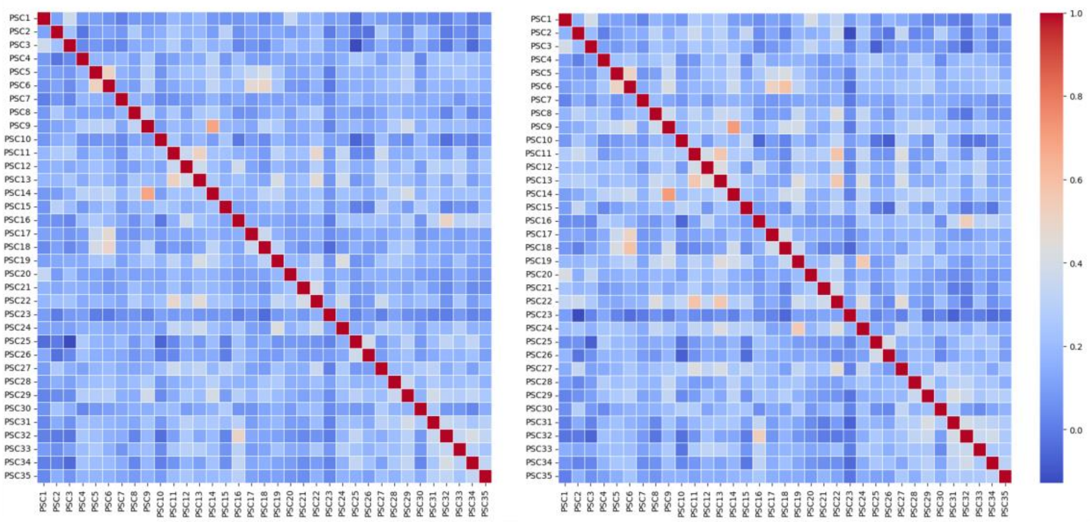

Figure 4.1: Heatmaps of the inter-item correlations of the parent-rated PSC administered to children (left) and adolescents (right).

Although majority of the correlations are on the weaker side and a few items of the PSC warrant further investigation in case of adolescents, the satisfactory magnitudes of Cronbach's alpha are sufficient to infer that the parent-rated PSC is a statistically reliable and consistent tool for screening of psychosocial dysfunction in Pakistani children and adolescents. Therefore, the proposed predictive models based on the parent-rated PSC will also be suitable as decision support systems for early detection and intervention of pediatric psychosocial dysfunction.

4.3 Hypothesis Testing

Independent samples t-tests performed between the means of results of children and adolescents are significant at the 0.05 level (Table 4.7). Therefore, the null hypothesis (H_0) is rejected, and it is concluded that a statistically significant difference exists between the results of children and adolescents. Additionally, the mean results of the PSC are higher in case of adolescents, indicating that the prevalence of psychosocial issues is higher among adolescents as compared to children.

Table 4.7: Results of independent samples t-test between the mean results of children and adolescents.

Results	**Descriptive Statistics**				**t-test for equality of means**	
	Children		**Adolescents**		**t-statistic**	**p-value**
	Mean	**SD**	**Mean**	**SD**		
Psychosocial Dysfunction	27.77	11.53	33.19	12.25	10.309	0.00
Attention Subscale	4.53	2.5	4.99	2.62	3.98	0.00
Internalizing Subscale	3.2	2.54	4.75	2.93	12.99	0.00
Externalizing Subscale	6.54	3.49	6.93	3.58	2.42	0.02

4.4 Feature Selection

Overall, both filter and wrapper methods reveal that majority, if not all, of the 36 predictive features are significant. Pearson Correlation between age and the target variable indicates that the correlation is significant at the 0.05 level (Table 4.8). However, the magnitudes of R for both age groups are less than 0.2, indicating a weak relationship. Therefore, the significance of age as a predictive feature of psychosocial dysfunction is investigated during model development.

Table 4.8: Pearson Correlation between age and the target variable.

Children		Adolescents	
R	**p-value**	**R**	**p-value**
0.097	0.00	0.186	0.00
R: Pearson Correlation Coefficient.			

According to the contingency coefficients (C) computed during Chi-square tests, item 23 of the PSC (Wants to be with you more than before) are weakly associated features for both children and adolescents, as their magnitude is within the 0.1 – 0.2 range (Table 4.9). Therefore, PSC23 is investigated as a significant predictive feature in subsequent steps. In conclusion, filter methods suggest a subset of 34 features, where age and PSC23 are removed.

Table 4.9: Chi-square tests between 35 items of the PSC and target variable.

Features	Children			Adolescents		
	χ^2	p-value	C	χ^2	p-value	C
PSC1	88.63	0.00	0.23	52.65	0.00	0.26
PSC2	138.15	0.00	0.28	76.36	0.00	0.31
PSC3	119.12	0.00	0.26	70.94	0.00	0.30
PSC4	238.47	0.00	0.36	102.90	0.00	0.36
PSC5	333.14	0.00	0.41	105.19	0.00	0.36
PSC6	294.24	0.00	0.39	139.45	0.00	0.40
PSC7	133.03	0.00	0.27	37.83	0.00	0.22
PSC8	211.51	0.00	0.34	125.05	0.00	0.39
PSC9	394.20	0.00	0.44	221.06	0.00	0.49
PSC10	94.89	0.00	0.23	44.44	0.00	0.24
PSC11	321.27	0.00	0.40	152.80	0.00	0.42
PSC12	306.05	0.00	0.40	138.41	0.00	0.40
PSC13	380.80	0.00	0.43	166.84	0.00	0.44
PSC14	455.14	0.00	0.46	202.31	0.00	0.47
PSC15	198.86	0.00	0.33	61.79	0.00	0.28
PSC16	246.12	0.00	0.36	86.64	0.00	0.33
PSC17	164.94	0.00	0.30	61.73	0.00	0.28
PSC18	204.98	0.00	0.33	74.23	0.00	0.31
PSC19	311.28	0.00	0.40	154.36	0.00	0.42
PSC20	153.57	0.00	0.29	64.27	0.00	0.29
PSC21	170.23	0.00	0.31	55.87	0.00	0.27
PSC22	323.76	0.00	0.40	133.98	0.00	0.40
PSC23	45.45	0.00	0.16	8.74	0.00	0.11
PSC24	252.87	0.00	0.36	133.40	0.00	0.40
PSC25	170.61	0.00	0.31	49.84	0.00	0.26
PSC26	147.36	0.00	0.29	40.35	0.00	0.23
PSC27	285.25	0.00	0.38	133.02	0.00	0.40
PSC28	283.35	0.00	0.38	106.62	0.00	0.36

PSC29	404.00	0.00	0.44	123.64	0.00	0.38
PSC30	152.53	0.00	0.29	87.10	0.00	0.33
PSC31	352.40	0.00	0.42	96.24	0.00	0.34
PSC32	251.86	0.00	0.36	68.46	0.00	0.30
PSC33	298.10	0.00	0.39	123.25	0.00	0.38
PSC34	202.97	0.00	0.33	79.06	0.00	0.32
PSC35	192.22	0.00	0.32	76.27	0.00	0.31

Results indicating a weak association are highlighted in red.
χ^2: Chi-square test statistic, C: Contingency Coefficient of Chi-square test.

The feature importances calculated during RFE are represented as bar charts in Figures 4.2 – 4.5 below. The features are arranged in descending order of their calculated Gini importance. These charts help in elucidation of the least significant features eliminated during RFE.

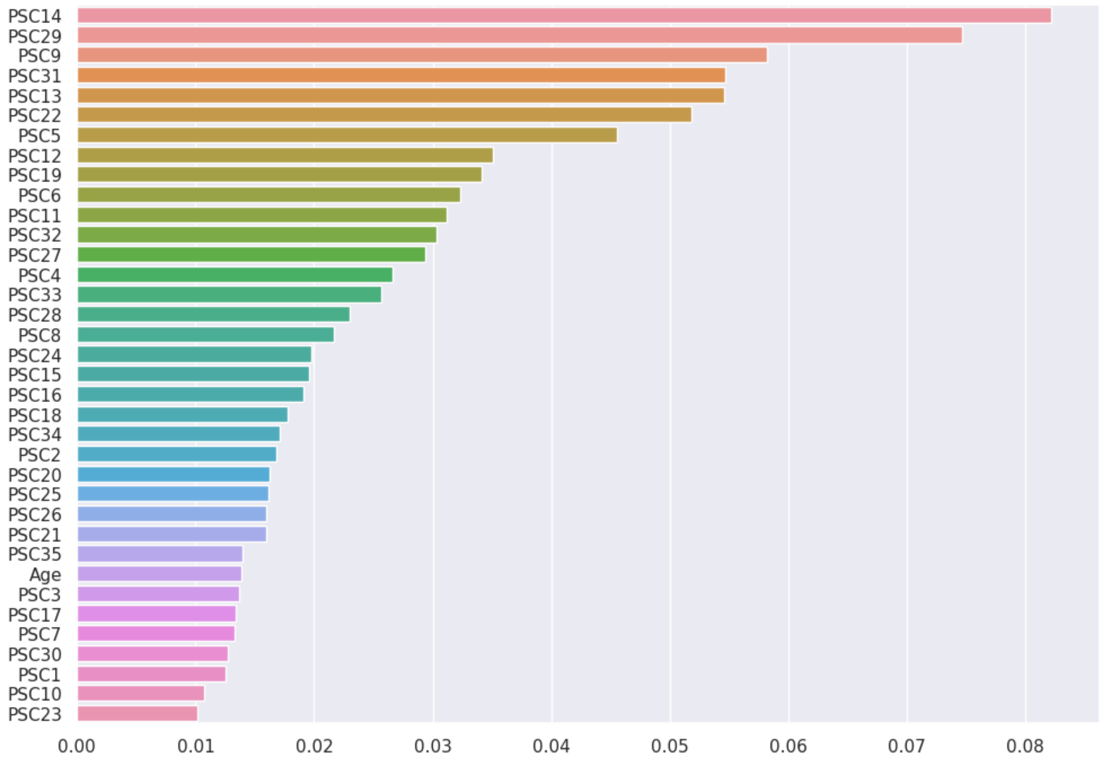

Figure 4.2: Feature Importances calculated for children during RF-RFE.

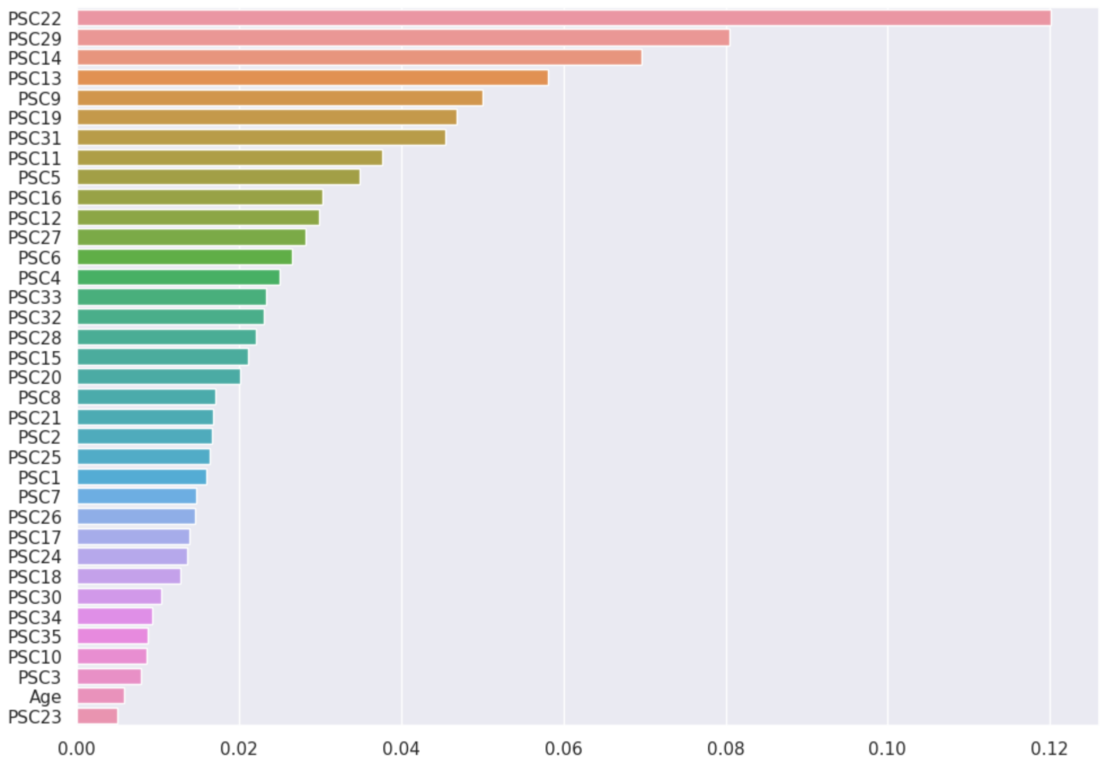

Figure 4.3: Feature Importances calculated for children during XGB-RFE.

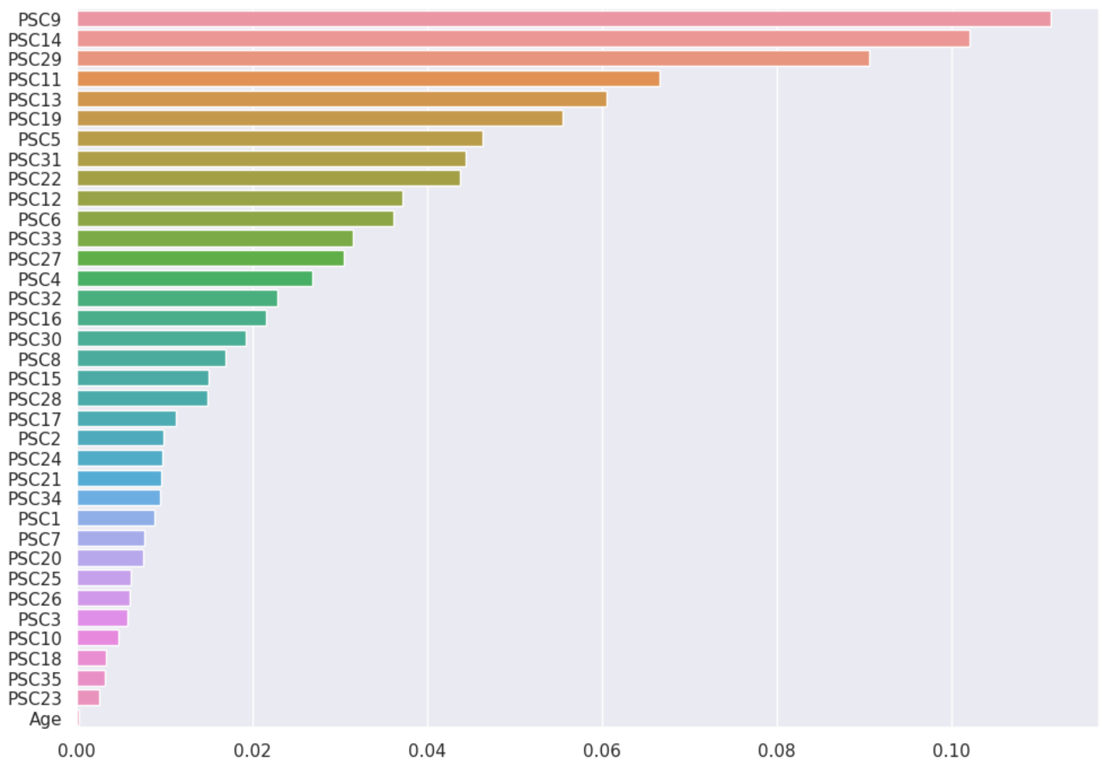

Figure 4.4: Feature Importances calculated for adolescents during RF-RFE.

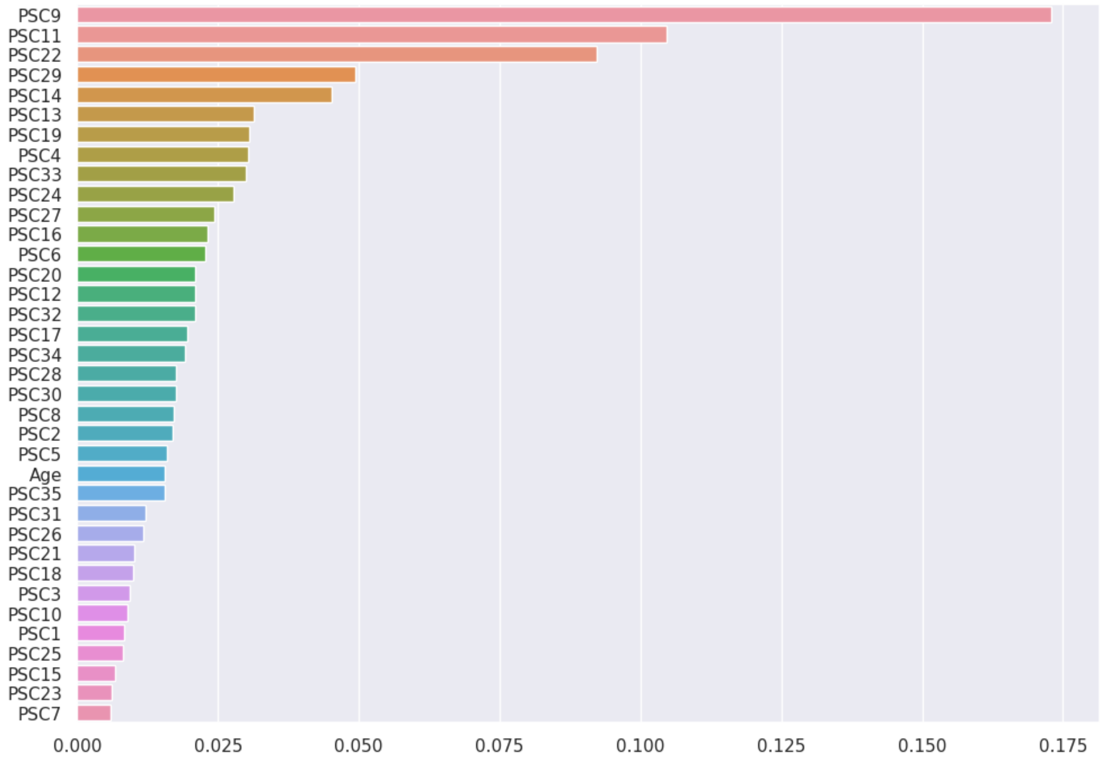

Figure 4.5: Feature Importances calculated for adolescents during XGB-RFE.

Number of optimal features providing the best accuracies during RFE for both age groups are defined in Table 4.10. For children, RF-RFE selects all 36 features to train a model with an accuracy of 0.94. This result is discarded and not considered during the feature selection process, as a subset of features is not produced. XGB-RFE provides the same result as filter methods where both age and PSC23 are eliminated and model accuracy is 0.938. Therefore, the results of filter and wrapper methods of feature selection are combined for children, as they are identical. The feature selection process yields only one subset of 34 features for children (Figure 4.6).

For adolescents, RF-RFE eliminates age and trains a model with an accuracy of 0.945. XGB-RFE removes item 7 of the PSC (Acts as if driven by a motor) and item 15 (Less interest in friends) in addition to item 23, resulting in a model with an accuracy of 0.933. Therefore, wrapper methods provide 2 subsets for adolescents, one with 34 features and the other with 33 features. Together with filter methods, the feature selection process for adolescents yields 3 different subsets (Figure 4.6).

Table 4.10: Results of Recursive Feature Elimination.

Age Group	RF-RFE			XGB-RFE		
	No. of features	**Eliminated features**	**Mean Accuracy**	**No. of features**	**Eliminated Features**	**Mean Accuracy**
Children	36	None	0.940	34	Age, PSC23	0.945
Adolescents	35	Age	0.938	33	PSC7, PSC15, PSC23	0.933

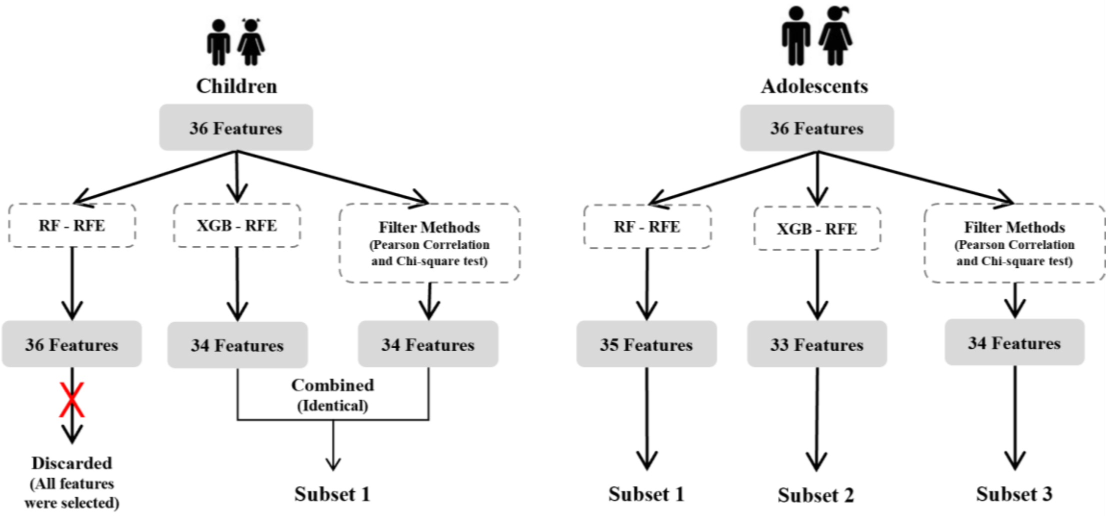

Figure 4.6: Flow diagram of the feature selection process.

Additionally, the top 10 significant features mutually selected by filter and wrapper methods have been analyzed to gain more insight regarding the most significant predictors of psychosocial dysfunction (Table 4.11). Although attention problems themselves are not prevalent in the study population (Table 4.3), items of the PSC concerned with attention problems are among the topmost significant features for both children and adolescents. Internalizing problems like hopelessness, sadness, self-esteem issues, and distress are more significant for adolescents as compared to children. This observation is compatible with published literature, as adolescence is the reported age of onset for internalizing problems like depression, anxiety, and self-esteem issues, making them more susceptible to these conditions [61]. Interestingly, items of the PSC concerned with externalizing problems like disobedience are exclusively identified as significant features in children. However, the risk of externalizing problems is reportedly higher in adolescents as compared to children [62]. These contrasting observations indicate that the presence of externalizing behaviors in children, which is uncharacteristic of their age, is a significant predictor of psychological impairment or dysfunction.

Table 4.11: Top 10 significant features selected mutually by filter and wrapper methods.

Children			Adolescents		
Feature	**Description**	**Category**	**Feature**	**Description**	**Category**
PSC14	Has trouble concentrating	A	PSC9	Distracted easily	A
PSC29	Does not listen to rules	E	PSC14	Has trouble concentrating	A
PSC9	Distracted easily	A	PSC13	Feels hopeless	I
PSC13	Feels hopeless	I	PSC11	Feels sad, unhappy	I
PSC31	Does not understand other people's feelings	E	PSC19	Is down on him or herself	I
PSC22	Worries a lot	I	PSC6	Less interested in school	O
PSC5	Has trouble with a teacher	O	PSC12	Is irritable, angry	O
PSC11	Feels sad, unhappy	I	PSC27	Seems to be having less fun	I
PSC19	Is down on him or herself	I	PSC22	Worries a lot	I
PSC12	Is irritable, angry	O	PSC24	Feels he or she is bad	O
Category defines which subscale each item belongs to on the PSC. A=Attention Subscale, I=Internalizing Subscale, E= Externalizing Subscale, O=Other Items which are not constituents of the 3 subscales.					

4.5 Model Development

For children, target classes are fairly balanced, with 855 Low-Risk instances (51.6%) and 802 High-Risk instances (48.4%). For adolescents, class imbalance is observed, with 260 Low-Risk instances (36.4%) and 455 High-Risk instances (63.6%).

4.5.1 Performance evaluation among different subsets of features

The accuracies of the ML algorithms trained on the original data set containing all features and the subsets are listed in Table 4.12. Mean accuracies of all models are also calculated to identify which subset of features results in the most accurate models overall. In case of children, the performance of SVM and LR remains constant with accuracies of 0.98. However, the accuracies of DT, RF, XGB, and MLP trained on subset 1 are higher, where the accuracy of DT increases from 0.79 to 0.8, and the accuracies of RF, XGB, and MLP increase from 0.92 to 0.94, 0.95, and 0.93, respectively. On average, subset 1 results in more accurate predictive models, as the mean accuracy of models trained on original data is 0.918 and the mean accuracy of subset 1 is 0.93. Therefore, this subset has been selected as the most significant set of features for prediction of psychosocial dysfunction in children, where age and PSC23 were removed.

For adolescents, XGB and MLP exhibit the same accuracy of 0.92 throughout, whereas the accuracies of DT, RF, SVM, and LR are the highest for subset 1. For DT, the accuracy is increased to 0.81 from 0.78. For RF, the accuracy increases slightly from 0.91 to 0.92. For SVM, it is raised from 0.97 to 0.98, and for LR, the accuracy is increased to 0.99 from 0.98. For the remaining 2 subsets, the accuracy either remains unchanged or decreases. The mean accuracy of models trained on original data is 0.913. For subset 1, it increases to 0.923, while for subset 2 and 3, it decreases to 0.905 and 0.907, respectively. As the mean accuracy of models trained on subset 1 is higher, it has been selected as the most significant set of features for prediction of psychosocial dysfunction in adolescents. In this subset, age was removed via RF-RFE.

Table 4.12: Accuracies of the models trained on all features of the data and subsets formed via feature selection.

Age Group	**Set of features**	**DT**	**RF**	**XGB**	**SVM**	**MLP**	**LR**	**Mean Accuracy**
Children	All (Original)	0.79	0.92	0.92	0.98	0.92	0.98	0.918
	Subset 1	**0.8**	**0.94**	**0.95**	**0.98**	**0.93**	**0.98**	**0.93**
Adolescents	All (Original)	0.78	0.91	0.92	0.97	0.92	0.98	0.913
	Subset 1	**0.81**	**0.92**	**0.92**	**0.98**	**0.92**	**0.99**	**0.923**
	Subset 2	0.78	0.91	0.92	0.95	0.92	0.95	0.905
	Subset 3	0.78	0.91	0.92	0.96	0.92	0.95	0.907
Subsets of features resulting in models with the best overall accuracies are highlighted in green.								

These findings evidence that age is not a significant predictive feature for the risk of psychosocial dysfunction among both children and adolescents. In similar studies, age has not been selected or defined as a significant feature as well [33],[34],[36],[38]. Interestingly, descriptive analysis (Table 4.3) and hypothesis testing (Table 4.7) suggest that the prevalence of psychosocial dysfunction varies between the two age groups, with the risk being higher in adolescents. This is in accordance with reported literature [61]. Moreover, studies pertaining to the prediction of suicidal behavior have recognized age as a significant feature [63],[64],[65]. These observations may be used to infer that the importance of age in the context of ML-based prediction of mental health issues is probably more important when considering specific behaviors like suicide. In the case of general psychosocial problems, age is not a significant predictive variable as evidenced by our research and similar studies.

Among children, item 23 of the PSC “Wants to be with you more than before” is also not a significant predictive feature. The same results have not been obtained for adolescents. This difference may be attributed to the contrast in the dependence of young children on their parents, and the growing independence of adolescents from their parents [66]. Younger children are physically and emotionally dependent on an older and more mature guardian. Therefore, a child wanting to spend more time with their parent may not indicate impaired mental health and well-being. However, adolescents begin to establish autonomy from their parents as they grow less dependent on an adult for their basic needs and develop a unique sense of individuality and identity [67]. Naturally, this causes them to become emotionally and physically distant. An adolescent seeking more emotional and physical closeness with their parents might indicate that they are experiencing some kind of psychological issue, as this behavior is contrary to the typical developmental trajectory of adolescence. As these findings align with the psychopathological perspective of child and adolescent development, the present study corroborates the efficacy of computational ML-driven approaches for the prediction of mental health issues.

4.5.2 Performance evaluation among algorithms

Detailed performance metrics including accuracy, precision, recall, and specificity of all six algorithms on the selected subsets of features are shown in Table 4.13. This shows that DT exhibits the poorest performance, RF, XGB, and MLP are mediocre, and SVM and LR perform exceptionally well, with performance metrics ranging between 0.97 and 0.99. In case of children, the performance metrics of both LR and SVM are identical, with an accuracy, precision, and recall of 0.98, and specificity of 0.99. However, analysis of the confusion matrix informs that LR is slightly better at the prediction of Low-Risk children (Table 4.14). Out of the 165 Low-Risk instances in the test set, SVM categorizes 163 of them accurately, while LR categorizes 164 accurately. While this is a very minute difference in performance, it is still appreciable for inferring that LR produces the most sensitive and specific predictive model for predicting psychosocial dysfunction in children aged 6 to 11 years.

For adolescents, the accuracy and precision of LR is slightly higher than SVM. The accuracy and precision of SVM are 0.98 and 0.97, respectively. The accuracy and precision of LR are slightly higher at 0.99 and 0.98, respectively. The recall and specificity are identical for both algorithms. The superior results of LR are also consolidated by the confusion matrix in Table 4.14, which indicates better performance of LR in terms of prediction of High-Risk adolescents. Out of the 96 High-Risk individuals in the test set, SVM recognizes 94 correctly while LR accurately categorizes 95 of them. Again, this is very minor difference in performance, but along with the better precision and accuracy, these results evidence that LR results in the best performing models for the prediction of psychosocial dysfunction in adolescents ages 12 to 17 years.

Table 4.13: Detailed performance metrics of the models trained on selected subsets of features.

Models	**Children**				**Adolescents**			
	Accuracy	**Precision**	**Recall**	**Specificity**	**Accuracy**	**Precision**	**Recall**	**Specificity**
DT	0.8	0.8	0.8	0.81	0.81	0.78	0.79	0.72
RF	0.94	0.94	0.94	0.96	0.92	0.89	0.9	0.87
XGB	0.95	0.95	0.95	0.98	0.92	0.89	0.91	0.91
SVM	**0.98**	**0.98**	**0.98**	**0.99**	**0.98**	**0.97**	**0.98**	**0.98**
MLP	0.93	0.93	0.93	0.93	0.92	0.89	0.91	0.88
LR	**0.98**	**0.98**	**0.98**	**0.99**	**0.99**	**0.98**	**0.98**	**0.98**
Models with the best performance metrics are highlighted in green.								

Table 4.14: Confusion Matrices of Support Vector Machine and Logistic Regression trained on the selected subsets of features.

Age Group	Models	Class labels		Predicted	
				Low-Risk	High-Risk
Children	SVM	Actual	Low-Risk	163	2
			High-Risk	5	162
	LR		Low-Risk	**164**	**1**
			High-Risk	5	162
Adolescents	SVM	Actual	Low-Risk	45	1
			High-Risk	2	94
	LR		Low-Risk	45	1
			High-Risk	**1**	**95**
Results that indicate better predictive performance are highlighted in green.					

In conclusion, Logistic Regression produces the more efficient predictive models for psychosocial dysfunction in children and adolescents, with accuracies of 0.98 and 0.99, respectively. Decision Tree exhibits the poorest performance with an accuracy of 0.8, while Random Forest, XGBoost, and Multilayer Perceptron are mediocre, with accuracies ranging between 0.92 to 0.95. These results are consistent with some reported literature that have utilized similar algorithms [38],[68],[69]. The optimal performance of SVM and LR can be attributed to the binary nature of our target variable. While multiclass versions of these algorithms exist, their performances as binary classifiers have been reported and commended frequently in literature [70], [71]. Specifically, LR is widely implemented for classification or regression in medicine and social science [60]. In addition to further supporting the efficiency of SVM and LR in binary classification and prediction problems, our findings also demonstrate that the implementation of more complex ensemble and deep learning methods like RF, XGB, and MLP may not be warranted for the development of AI-driven processes for detection of mental health issues, as far as binary target variables are concerned.

5. CONCLUSIONS AND FUTURE RECOMMENDATIONS

5.1 Key Findings and Strengths

The main findings of this study indicate that AI-driven predictive models can efficiently screen Pakistani children and adolescents for psychosocial dysfunction at accuracy levels of 98 to 99 percent. Out of the six selected algorithms, SVM and LR are the top-performing models, where LR is slightly more accurate, specific, and sensitive for the prediction of low-risk children and high-risk adolescents. Feature selection methods reveal that age is not a significant predictive feature in the context of AI-driven prediction of both children and adolescents, and item 23 (Wants to spend more time with you than before) of the PSC is not a significant feature for children only. Attention problems and internalizing problems are among the top 10 significant features for the prediction of psychosocial dysfunction in both age groups. In case of children, externalizing problems are also included in these topmost significant features. These observations further indicate differences between the psychopathology of children and adolescents. As most of our findings resonate with published literature regarding child and adolescent mental health and ML-driven prediction of pediatric mental health issues, the present study further evidences the efficiency of AI in mental health informatics.

To the best of our knowledge, the study is the first of its kind in both aspects of utilizing the PSC and local data of Pakistani children and adolescents for the development of AI-driven predictive models. This novel contribution within the field of mental health informatics can promote further research and development regarding the integration of AI in mental healthcare practices specifically in LMICs, where research is limited in this context and the burden of impaired mental health and well-being is exacerbated by various challenges like societal stigma, inadequate access to resources, and financial strain of mental health consultations.

5.2 Limitations and Future Recommendations

Practically, the proposed predictive models can be implemented in hospitals, clinics, schools, or even used at home as they are based on the free parent-rated version of the PSC. Within hospitals, the models can be deployed in pediatric wards in addition to psychiatry departments, as the influx of young patients is generally higher for pediatricians as opposed to child and adolescent psychiatrists. However, optimization of the models in light of the following identified limitations are prerequisites to ensure the success of their implementation.

While the study has utilized diverse data of Pakistani children and adolescents residing in 89 different cities across the country, the sample size is still limited and not generalizable for the entire population under the age of 18 in Pakistan, as 2,372 individuals are not representative for a group of 212 million [8]. A larger cohort will improve the generalizability of the predictive models. Techniques of synthetic data generation may also be incorporated in future studies to increase the sample size [72]. These can also address the issue of class imbalance, as is observed in the proportions of Low-Risk and High-Risk adolescents in our study population.

Moreover, the predictive features for psychosocial dysfunction were limited to age and the 35 items of the PSC. Socioeconomic, environmental, and other clinical data of the participants such as medical history has not been considered, all of which have been proven to have significant implications on a young individual's well-being [17]. In the future, these aspects should be incorporated to improve the efficacy and precision of the proposed AI-driven predictive models.

In the present study, the predictive models have been limited to psychosocial dysfunction only. The subscales of the PSC have not been incorporated, as the interpretation of such a comprehensive multiclass predictive model would have been quite complex and beyond the scope of the study. Nevertheless, the current findings can be investigated and expanded upon from this perspective in the future for the development of predictive models that screen for attention, internalizing, and externalizing problems in addition to overall psychosocial dysfunction.

REFERENCES

[1] D. Yurgelun-Todd, “Emotional and cognitive changes during adolescence,” *Curr Opin Neurobiol*, vol. 17, no. 2, pp. 251–257, 2007, doi: https://doi.org/10.1016/j.conb.2007.03.009.

[2] K. von Klitzing, M. Döhnert, M. Kroll, and M. Grube, “Mental disorders in early childhood,” *Dtsch Arztebl Int*, vol. 112, no. 21–22, p. 375, 2015.

[3] W. E. Copeland, D. Wolke, L. Shanahan, and E. J. Costello, “Adult Functional Outcomes of Common Childhood Psychiatric Problems: A Prospective, Longitudinal Study,” *JAMA Psychiatry*, vol. 72, no. 9, pp. 892–899, Sep. 2015, doi: 10.1001/jamapsychiatry.2015.0730.

[4] C. Kieling *et al.*, “Child and adolescent mental health worldwide: evidence for action,” *The Lancet*, vol. 378, no. 9801, pp. 1515–1525, 2011, doi: https://doi.org/10.1016/S0140-6736(11)60827-1.

[5] S. Choudhury, S.-J. Blakemore, and T. Charman, “Social cognitive development during adolescence,” *Soc Cogn Affect Neurosci*, vol. 1, no. 3, p. 165, 2006.

[6] N. Imran, I. Aamer, M. I. Sharif, Z. H. Bodla, and S. Naveed, “Psychological burden of quarantine in children and adolescents: A rapid systematic review and proposed solutions,” *Pak J Med Sci*, vol. 36, no. 5, p. 1106, 2020.

[7] A. Javed, M. N. S. Khan, A. Nasar, and A. Rasheed, “Mental healthcare in Pakistan,” *Taiwanese Journal of Psychiatry*, vol. 34, no. 1, p. 6, 2020.

[8] United Nations Children’s Fund, “Annual Report 2020: An in-depth overview of UNICEF Pakistan’s activities in 2020,” Islamabad, Pakistan, Apr. 2021. Accessed: Nov. 24, 2023. [Online]. Available: https://www.unicef.org/pakistan/reports/annual-report-2020

[9] A. Khalid, F. Qadir, S. W. Y. Chan, and M. Schwannauer, “Adolescents’ mental health and well-being in developing countries: a cross-sectional survey from Pakistan,” *Journal of Mental Health*, vol. 28, no. 4, pp. 389–396, Jul. 2019, doi: 10.1080/09638237.2018.1521919.

[10] S. Farooq, T. Yousaf, and S. Shahzad, “PREVALENCE OF EMOTIONAL AND BEHAVIOURAL PROBLEMS AMONG ADOLESCENTS IN PAKISTAN: A CROSS-SECTIONAL STUDY,” *J Pak Psychiatr Soc*, vol. 20, no. 01, 2023.

[11] E. U. Syed, S. A. Hussein, and S. Mahmud, "Screening for emotional and behavioural problems amongst 5–11-year-old school children in Karachi, Pakistan," *Soc Psychiatry Psychiatr Epidemiol*, vol. 42, no. 5, pp. 421–427, 2007, doi: 10.1007/s00127-007-0188-x.

[12] World Health Organization, "Mental health atlas 2017. Geneva: World Health Organization; 2018," *World Health Organization WHO MiNDbank.* Accessed: Jul. 21, 2023.

[13] M. O. Husain *et al.*, "Stigma toward mental and physical illness: attitudes of healthcare professionals, healthcare students and the general public in Pakistan," *BJPsych Open*, vol. 6, no. 5, p. e81, 2020.

[14] J. Goldman, "Mental Health Screening for Children & Teens," Medical Home Portal. Accessed: Mar. 16, 2024. [Online]. Available: https://www.medicalhomeportal.org/clinical-practice/screening-and-prevention/mental-health-screening-for-children-and-teens

[15] M. S. Jellinek, J. M. Murphy, J. Robinson, A. Feins, S. Lamb, and T. Fenton, "Pediatric Symptom Checklist: Screening school-age children for psychosocial dysfunction," *J Pediatr*, vol. 112, no. 2, pp. 201–209, 1988, doi: https://doi.org/10.1016/S0022-3476(88)80056-8.

[16] "Pediatric Symptom Checklist," Massachusetts General Hospital. Accessed: Mar. 17, 2024.

[17] I. van Kamp *et al.*, "Early environmental quality and life-course mental health effects: The Equal-Life project," *Environmental Epidemiology*, vol. 6, no. 1, 2022.

[18] S. S. Montani Manuel, "Artificial Intelligence in Clinical Decision Support: a Focused Literature Survey," *Yearb Med Inform*, vol. 28, no. 01, pp. 120–127, 2019, doi: 10.1055/s-0039-1677911.

[19] B. Bista, P. Thapa, D. Sapkota, S. B. Singh, and P. K. Pokharel, "Psychosocial problems among adolescent students: an exploratory study in the central region of Nepal," *Front Public Health*, vol. 4, p. 158, 2016.

[20] E. S. Soliman, R. S. Mahdy, H. A. Fouad, R. A. Abbas, and A. Fayed, "Multiple risk factors affecting childhood psychosocial dysfunction in primary school Egyptian children," *Middle East Current Psychiatry*, vol. 27, no. 1, pp. 1–9, 2020.

[21] L. A. Clark, H. Nuzum, and E. Ro, "Manifestations of personality impairment severity: comorbidity, course/prognosis, psychosocial dysfunction, and 'borderline' personality features," *Curr Opin Psychol*, vol. 21, pp. 117–121, 2018, doi: https://doi.org/10.1016/j.copsyc.2017.12.004.

[22] N. Eskander, “The psychosocial outcome of conduct and oppositional defiant disorder in children with attention deficit hyperactivity disorder,” *Cureus*, vol. 12, no. 8, 2020.

[23] M. J. Knight and B. T. Baune, “The direct and indirect relationship between social cognition and psychosocial dysfunction in major depressive disorder,” *Front Psychiatry*, vol. 10, p. 449654, 2019.

[24] N. P. Friedman *et al.*, “Greater Attention Problems During Childhood Predict Poorer Executive Functioning in Late Adolescence,” *Psychol Sci*, vol. 18, no. 10, pp. 893–900, Oct. 2007, doi: 10.1111/j.1467-9280.2007.01997.x.

[25] D. Cicchetti and S. L. Toth, “A developmental perspective on internalizing and externalizing disorders,” *Internalizing and externalizing expression of dysfunction*, vol. 2, pp. 1–19, 2014.

[26] E. M. Parker, J. Jacobson, M. D. Pullmann, and S. E. U. Kerns, “Identifying Psychosocial Problems Among Children and Youth in the Child Welfare System Using the PSC-17: Exploring Convergent and Discriminant Validity with Multiple Informants,” *Child Psychiatry Hum Dev*, vol. 50, no. 1, pp. 108–120, 2019, doi: 10.1007/s10578-018-0824-6.

[27] H. Fayyaz, T.-L. T. Phan, H. T. Bunnell, and R. Beheshti, “Predicting attrition patterns from pediatric weight management programs,” in *Machine Learning for Health*, PMLR, 2022, pp. 326–342.

[28] N. K. Iyortsuun, S.-H. Kim, M. Jhon, H.-J. Yang, and S. Pant, “A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis,” in *Healthcare*, MDPI, 2023, p. 285.

[29] I. El Naqa and M. J. Murphy, *What is machine learning?* Springer, 2015.

[30] V. Nasteski, “An overview of the supervised machine learning methods,” *Horizons. b*, vol. 4, no. 51–62, p. 56, 2017.

[31] F. Hahne, W. Huber, R. Gentleman, S. Falcon, R. Gentleman, and V. J. Carey, “Unsupervised machine learning,” *Bioconductor case studies*, pp. 137–157, 2008.

[32] N. Zhang *et al.*, “Prediction of adolescent subjective well-being: A machine learning approach,” *Gen Psychiatr*, vol. 32, no. 5, p. e100096, Sep. 2019, doi: 10.1136/gpsych-2019-100096.

[33] A. E. Tate, R. C. McCabe, H. Larsson, S. Lundström, P. Lichtenstein, and R. Kuja-Halkola, “Predicting mental health problems in adolescence using machine learning techniques,” *PLoS One*, vol. 15, no. 4, pp. e0230389-, Apr. 2020, [Online]. Available: https://doi.org/10.1371/journal.pone.0230389

[34] R. Qasrawi *et al.*, “Machine learning techniques for identifying mental health risk factor associated with schoolchildren cognitive ability living in politically violent environments,” *Front Psychiatry*, vol. 14, p. 1071622, 2023.

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

[36] K. L. H. Carpenter, P. Sprechmann, R. Calderbank, G. Sapiro, and H. L. Egger, “Quantifying Risk for Anxiety Disorders in Preschool Children: A Machine Learning Approach,” *PLoS One*, vol. 11, no. 11, pp. e0165524-, Nov. 2016, [Online]. Available: https://doi.org/10.1371/journal.pone.0165524

[37] F. Ge, Y. Li, M. Yuan, J. Zhang, and W. Zhang, “Identifying predictors of probable posttraumatic stress disorder in children and adolescents with earthquake exposure: a longitudinal study using a machine learning approach,” *J Affect Disord*, vol. 264, pp. 483–493, 2020.

[38] R. Qasrawi, S. P. V. Polo, D. A. Al-Halawa, S. Hallaq, and Z. Abdeen, “Assessment and prediction of depression and anxiety risk factors in schoolchildren: machine learning techniques performance analysis,” *JMIR Form Res*, vol. 6, no. 8, p. e32736, 2022.

[39] L. Godoy, A. S. Carter, R. B. Silver, S. Dickstein, and R. Seifer, “Infants and Toddlers Left Behind: Mental Health Screening and Consultation in Primary Care,” *Journal of Developmental & Behavioral Pediatrics*, vol. 35, no. 5, 2014, [Online]. Available: https://journals.lww.com/jrnldbp/fulltext/2014/06000/infants_and_toddlers_left_behind__mental_health.4.aspx

[40] “SehatYab - Online Psychiatrist / Psychologist in Pakistan.” Accessed: Jul. 21, 2023. [Online]. Available: https://www.sehatyab.com/

[41] IBM Corp, “IBM SPSS for Windows.” IBM Corp, Armonk, NY, 2011.

[42] G. VanRossum and F. L. Drake, *The python language reference*, vol. 561. Python Software Foundation Amsterdam, Netherlands, 2010.

[43] E. Cohen Hubal *et al.*, “Identifying important life stages for monitoring and assessing risks from exposures to environmental contaminants: Results of a World Health Organization review,” *Regul Toxicol Pharmacol*, vol. 69, Oct. 2013, doi: 10.1016/j.yrtph.2013.09.008.

[44] S. U. Hamdani, H. Javed, A. Warraitch, A. Rahman, A. T. Nizami, and F. A. Minhas, “Prevalence of psychosocial distress in school going adolescents in rural Pakistan: findings from a cross-sectional epidemiological survey,” *BJPsych Open*, vol. 7, no. S1, pp. S56–S57, 2021.

[45] D. G. Bonett and T. A. Wright, "Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning," *J Organ Behav*, vol. 36, no. 1, pp. 3–15, 2015.

[46] M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha," *Int J Med Educ*, vol. 2, p. 53, 2011.

[47] B. Damasceno and B. Damasceno, "Hypothesis Testing," *Research on Cognition Disorders: Theoretical and Methodological Issues*, pp. 157–166, 2020.

[48] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, and M. Lang, "Benchmark for filter methods for feature selection in high-dimensional classification data," *Comput Stat Data Anal*, vol. 143, p. 106839, 2020, doi: https://doi.org/10.1016/j.csda.2019.106839.

[49] P. Sedgwick, "Pearson's correlation coefficient," *Bmj*, vol. 345, 2012.

[50] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: appropriate use and interpretation," *Anesth Analg*, vol. 126, no. 5, pp. 1763–1768, 2018.

[51] N. Pandis, "The chi-square test," *American journal of orthodontics and dentofacial orthopedics*, vol. 150, no. 5, pp. 898–899, 2016.

[52] S.-K. Hong, H. Kim, S. Lee, and Y.-S. Moon, "Secure multiparty computation of chi-square test statistics and contingency coefficients," in *2017 ieee 3rd international conference on big data security on cloud (bigdatasecurity), ieee international conference on high performance and smart computing (hpsc), and ieee international conference on intelligent data and security (ids)*, IEEE, 2017, pp. 53–56.

[53] N. Nnamoko, F. Arshad, D. England, J. Vora, and J. Norman, "Evaluation of filter and wrapper methods for feature selection in supervised machine learning," *Age (Omaha)*, vol. 21, no. 81, pp. 32–33, 2014.

[54] B. F. Darst, K. C. Malecki, and C. D. Engelman, "Using recursive feature elimination in random forest to account for correlated variables in high dimensional data," *BMC Genet*, vol. 19, pp. 1–6, 2018.

[55] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[56] G. Bonaccorso, *Machine learning algorithms*. Packt Publishing Ltd, 2017.

[57] A. Cutler, D. R. Cutler, and J. R. Stevens, "Random forests," *Ensemble machine learning: Methods and applications*, pp. 157–175, 2012.

[58] T. Chen *et al.*, “Xgboost: extreme gradient boosting,” *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.

[59] M.-C. Popescu, V. E. Balas, L. Perescu-Popescu, and N. Mastorakis, “Multilayer perceptron and neural networks,” *WSEAS Transactions on Circuits and Systems*, vol. 8, no. 7, pp. 579–588, 2009.

[60] T. G. Nick and K. M. Campbell, “Logistic Regression,” in *Topics in Biostatistics*, W. T. Ambrosius, Ed., Totowa, NJ: Humana Press, 2007, pp. 273–301. doi: 10.1007/978-1-59745-530-5_14.

[61] R. C. Kessler, G. P. Amminger, S. Aguilar-Gaxiola, J. Alonso, S. Lee, and T. B. Üstün, “Age of onset of mental disorders: a review of recent literature,” *Curr Opin Psychiatry*, vol. 20, no. 4, 2007, [Online]. Available: https://journals.lww.com/co-psychiatry/fulltext/2007/07000/age_of_onset_of_mental_disorders__a_review_of.10.aspx

[62] J. E. Bates, K. A. Dodge, J. E. Lansford, I. T. Petersen, and G. S. Pettit, “Describing and predicting developmental profiles of externalizing problems from childhood to adulthood,” *Dev Psychopathol*, vol. 27, no. 3, pp. 791–818, 2015, doi: DOI: 10.1017/S0954579414000789.

[63] C. G. Walsh, J. D. Ribeiro, and J. C. Franklin, “Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning,” *Journal of child psychology and psychiatry*, vol. 59, no. 12, pp. 1261–1270, 2018.

[64] M. C. Navarro *et al.*, “Machine learning assessment of early life factors predicting suicide attempt in adolescence or young adulthood,” *JAMA Netw Open*, vol. 4, no. 3, pp. e211450–e211450, 2021.

[65] C. Su, R. Aseltine, R. Doshi, K. Chen, S. C. Rogers, and F. Wang, “Machine learning for suicide risk prediction in children and adolescents with electronic health records,” *Transl Psychiatry*, vol. 10, no. 1, p. 413, 2020.

[66] Ikromova Sitora Akbarovna, “SOCIO-PSYCHOLOGICAL FACTORS OF BEHAVIOR FORMATION IN ADOLESCENTS.,” *TECHNICAL SCIENCE RESEARCH IN UZBEKISTAN*, vol. 1, no. 5, pp. 184–191, Dec. 2023, [Online]. Available: https://universalpublishings.com/index.php/tsru/article/view/3324

[67] M. Luciana, “Adolescent brain development in normality and psychopathology,” *Dev Psychopathol*, vol. 25, no. 4pt2, pp. 1325–1345, 2013, doi: DOI: 10.1017/S0954579413000643.

[68] A. E. Tate, R. C. McCabe, H. Larsson, S. Lundström, P. Lichtenstein, and R. Kuja-Halkola, “Predicting mental health problems in adolescence using machine learning techniques,” *PLoS One*, vol. 15, no. 4, p. e0230389, 2020.

[69] K. Malaquias, T. Lima, R. Santana, F. Salgado, M. Teodoro, and C. Nobre, "Classification and characterization of children and adolescents with depressive symptomatology using machine learning," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, IEEE, 2019, pp. 534–539.

[70] A. B. Musa, "Comparative study on classification performance between support vector machine and logistic regression," *International Journal of Machine Learning and Cybernetics*, vol. 4, no. 1, pp. 13–24, 2013, doi: 10.1007/s13042-012-0068-x.

[71] M. Awad and R. Khanna, "Support Vector Machines for Classification," in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, M. Awad and R. Khanna, Eds., Berkeley, CA: Apress, 2015, pp. 39–66. doi: 10.1007/978-1-4302-5990-9_3.

[72] H. Murtaza, M. Ahmed, N. F. Khan, G. Murtaza, S. Zafar, and A. Bano, "Synthetic data generation: State of the art in health care domain," *Comput Sci Rev*, vol. 48, p. 100546, 2023, doi: https://doi.org/10.1016/j.cosrev.2023.100546.