Predicting suicide risks using Machine Learning among medical students in Bangladesh

By

Zulker Nien
ID: 012221045

Submitted in Partial Fulfillment of the Requirements For the Degree of Master of Science in Computer Science and Engineering

September 20,2024

Department of Computer Science and Engineering
United International University

Approval Certificate

This thesis titled "Predicting suicide risks using Machine Learning among medical students in Bangladesh" submitted by **Zulker Nien**, Student ID: **012221045**, has been accepted as Satisfactory in fulfillment of the requirement for the degree of Master of Science in Computer Science and Engineering on 31st December, 2024.

Board of Examiners

1. Supervisor

____

Dr. Khondaker Abdullah Al Mamun
Professor & Director - AIMS Lab
Department of CSE
United International University
Dhaka, Bangladesh

2. Head Examiner

____

Prof. Dr. Mohammad Nurul Huda
Professor and Head,
Department of CSE
United International University
Dhaka, Bangladesh

3. Examiner-I

____

Dr. Md. Saddam Hossain Mukta
Associate Professor
Department of CSE
United International University
Dhaka, Bangladesh

4. Examiner-II

____

Prof. Dr. Muhammad Nomani Kabir
Professor and Undergraduate Program Coordinator,
Department of CSE
United International University
Dhaka, Bangladesh

5. Ex-Officio

____

Prof. Dr. Md. Motaharul Islam
Professor and Director of MSCSE
Department of CSE
United International University
Dhaka, Bangladesh

Declaration

This is to certify that the work entitled **"Predicting suicide risks using Machine Learning among medical students in Bangladesh"** is the outcome of the research carried out by me under the supervision of Professor Dr. Khondaker Abdullah Al Mamun.

Zulker Nien
Student ID: 012221045
Department of CSE
United International University
Dhaka, Bangladesh

In my capacity as supervisor of the candidate's thesis, I certify that the above statements are true to the best of my knowledge.

Dr. Khondaker Abdullah Al Mamun
Professor & Director - AIMS Lab
Department of CSE
United International University
Dhaka, Bangladesh

Abstract

Medical students face multiple challenges that can significantly impact their mental health. The intense academic pressure, exposure to death and long working hours make them vulnerable to stress, anxiety and sometimes depression. In Low and middle-income countries (LMIC), additional factors like social stigma, peer pressure and family issues like parental expectation can further contribute to their mental instability, ultimately leading to the ideation of suicide. It is found that in Bangladesh alone, the prevalence of suicidal ideation was 27.4% which is higher than the global prevalence. Even though the rates keep increasing, early signs of suicidal ideation go undiagnosed due to various reasons such as social and cultural stigma. Thus, this thesis brings forth a model developed for predicting suicidal risks among medical students in Bangladesh due to the complexity of this topic. It aims to contribute to the knowledge about mental health by creating a feasible machine learning model to identify suicidal thoughts and access to early intervention among Medical Students in Bangladesh. Three publicly available datasets were selected where multiple types of data such as symptoms, demographic & scale data were utilized to extract information through surveys. Four machine learning algorithms were employed on each dataset, after careful cleaning and preprocessing for all the models. For feature extraction, LASSO, Pearson's Correlation Coefficient (PCC), Sequential Forward Selection (SFS), and Weighted Sum Feature Selection (WSFS) were utilized. When Random Forest was applied to the features selected by WSFS on Dataset A, it proved to be the best model with an accuracy of 0.97 and an AUC of 0.99, further supported by Cohen's Kappa value of 0.95. LASSO selected features from all datasets performed well, and algorithms gave good results as well. Overall, the Random Forest performed better than Decision Trees, while Logistic Regression and SVM gave almost similar results. The combined dataset model, however, was not able to showcase good performance.

Acknowledgements

First and foremost, I would like to begin with thanking the owner of all thanks, Allah, for providing me with the opportunity to accumulate knowledge and encounter with all the people supporting me in completing this thesis work.

Next, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Khondaker Abdullah Al Mamun, for his continuous encouragement, remarks, and patience. His support influenced me to explore knowledge more efficiently and become as good as a researcher I have the potential to be. I would also like to extend my thanks to Dr. Marzia Zaman for helping me analyze the datasets I worked on and provide expert feedback.

I am deeply thankful to my wife, who also happens to be my classmate, study partner, reviewer and my better half. Without her moral support, I could not have moved ahead with this research. Her constant push, encouragement and insightful knowledge is something that I will forever cherish. I would also like to extend my thanks to my parents, parents-in-law, and my friend's family who took care of my child on separate occasions while I attended classes with my wife and worked on our individual thesis. Their faith in me has kept my motivation high during all this time and allowed me to continue my education.

Furthermore, I would like to extend my thanks to the research assistants from IRIIC who helped me with their expertise in times of need. I also want to acknowledge the advancements in technology, including the internet and artificial intelligence tools for research, which have enhanced access to knowledge and significantly supported this work.

Publication List

The main contributions of this research are either published or accepted or in preparation in journals and conferences as mentioned in the following list:

Journal Articles

1. The Use of Artificial Intelligence in Advancing Mental Health Interventions: A Systematic Review"

   Submitted in Heliyon (Q1 Journal)

   Manuscript number: **HELIYON-D-24-67395R1**

2. Zulker Nien, Khondaker A. Mamun "Predicting suicide risks using Machine Learning among medical students in Bangladesh"

   Journal in preparation

Additional Publications

Following is the list of relevant publications published in the course of the research that is not included in the thesis:

1. Screening for Common Cancers

   (Published by Elsevier)

   Book Title: **Artificial Intelligence in e-Health Framework, Volume 1**

2. The Use of Artificial Intelligence in Advancing Sexual and Reproductive Health and Rights (SRHR) Interventions: A Systematic Review

   Submitted in Heliyon (Q1 Journal)

   Manuscript number: **HELIYON-D-24-70207**

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter explores the overall problem of mental health challenges worldwide. The main focus is in Bangladesh and other low and middle income countries, where resources are limited despite increasing needs. The chapter funnels down from broad mental health statistics to the specific issue of suicide risk among medical students in Bangladesh. Moreover, this chapter highlights why traditional assessment approaches fall short and introduces how computational intelligence might offer new pathways for early identification of at-risk individuals.

1.1 Background

Mental health problems are on the rise worldwide due to factors such as family issues, stressful life events, substance use and other common issues [1]. It is a complex and often stigmatizing condition that can affect the size of one's thoughts, feelings, and actions. Approximately 13% of the global population (971 million people) suffer from some form of mental illness [2]. Among them, the number of individuals with a serious mental illness has been observed to significantly rise among university students over the past few years [3]. Globally, one-fifth of university students suffer from a mental disorder in every given year [4].

When examining student mental health, it is essential to note the underlying dependencies associated with their situation. One study found that students dealing with mental disorders typically face challenges such as diminished self-esteem, social stigma, instances of bullying, and difficulties in both family and social relationships [5]. Statistics on depression show a concerning trend of rising depression among 17-year-olds, emphasizing the importance of collegiate mental health services to address this issue as they transition into college [6]. Furthermore, one-third of first-year college students screened positive for at least one mental disorder, highlighting the need for mental health services on campuses [7]. Although many studies have been conducted on the mental health of students worldwide, studies remain relatively low for low- and middle-income countries (LMICs). A study conducted in Vietnam points towards the scarcity of mental health resources in LMICs [8]. Another study led by the World Health Organization (WHO) found that the majority of populations in LMICs do not have access to effective psychological and pharmacological interventions [9]. Bangladesh is yet another LMIC that is facing a rise in mental health problems and is considering implementing community-based mental health services to address the issue [10]. However, individuals with mental health concerns in this demography face social stigma and lack emotional support [11]. Also, the healthcare system in Bangladesh is struggling with a shortage of skilled mental health professionals. This increases the demand for additional trained experts to deliver sufficient care and services, posing a challenge in delivering timely and effective support to all individuals who may require it [12]. Furthermore, the lack of mental health awareness and acknowledgement in this country, along with the cultural stigma surrounding mental health, creates a significant barrier to addressing and effectively supporting the mental well-being of the Bangladeshi citizens [13]. Although, one study contradicts that mental health awareness is high among Bangladeshi university students [14], knowledge on mental health is insufficient. This inadequacy contributes to the widespread prevalence of mental health issues in Bangladesh, including depression, anxiety, study-related problems, trauma, abuse, etc.

One such issue is suicidal ideation which is a deeply concerning aspect of mental health, even though it is a neglected public health concern in Bangladesh, with an average suicide rate of 39.6/100000 population/year [15]. This rate of suicide in Bangladesh increased between 2002 and 2015, particularly among adolescents, females, and urban dwellers [16]. Students in particular were the most vulnerable to suicide amidst the vast majority, with academic stress and relationship concerns being common factors [17]. A 2022 cross-sectional study of Bangladeshi university students found that 13.4% had experienced suicidal thoughts in the past year, 6% had made suicide plans and 4.4% had attempted suicide at some point in their life [18].

1.2 Motivation

Enhancing the mental well-being of individuals who view suicide as a way to escape emotional distress requires the identification of such feelings and struggles. The traditional method of identifying suicide risk can be achieved through the use of scales such as suicide risk assessment (SRA) and suicide risk formulation (SRF) [19]. Other methods include a one-dimensional stratification system based on severity [20] or explicit communications of suicidal ideation [21]. However, these methods may rely on self-reporting or assessments by mental health professionals, which can be subjective and may not capture the full extent of an individual's risk [22]. Furthermore, traditional methods may not adequately capture the complexity and dynamic nature of suicide risk, as they rely on observable and reported symptoms, behaviors, and historical factors [23].

Artificial Intelligence (AI) can be a useful tool to predict the outbreaks of suicide and identify high-risk individuals [24]. It refers to the development of computer systems that can perform tasks that typically require human intelligence, such as visual perception, speech recognition, decision-making, and problem-solving [25]. By combining large amounts of data with fast processing and smart algorithms, AI allows a program to learn from patterns or features in the data automatically [26]. By leveraging the use of AI, further contribution to the early detection of suicide among medical students in Bangladesh can be achieved.

1.3 Objectives

This study focuses on predicting the risk of suicide among medical students of Bangladesh by using sociodemographic, symptoms, and scale data with machine learning for quick assessment of their mental health. The objectives of this research are as follows:

- To reduce diagnosis delay, the use of this suicide prediction model inside medical schools in a scheduled manner can help identify at risk individuals
- To Improve access to care, this machine learning model will help to bridge the gap between at risk medical students and mental healthcare providers.

- To Assist with resource allocation, experienced psychiatrists can be referred to high-risk individuals, ensuring that the mental health resources are efficiently allocated and that high-risk individuals receive timely, personalized care for effective intervention.

1.4 Outline of thesis

The subsequent chapters of this thesis are organized as follows: Chapter II provides a literature review summarizing previous studies on predicting risks of suicide among Bangladeshi students. In Chapter III, the data associated with suicide prediction particularly among medical students are analyzed and some similar machine learning suicide prediction models are discussed. In Chapter IV, the methodology behind the predictive model is introduced. Moving onward to Section V, the results and evaluation of model performance are analyzed. Lastly, in Chapter VI, a comprehensive discussion and analysis of the findings are provided, placing them in a broader context to explore their implications.

Chapter 2

Literature Review

Mental health, particularly within the academic community, has been one of the major topics of discussion in recent years. This is due to the stress that students have to endure to meet education and training demands. Among this vast population of students are medical students, who stand out to be more vulnerable to mental health challenges [27]. Factors contributing to their vulnerability include intense academic pressure, financial burdens, sleep deprivation, exposure to sickness and death, and mistreatment during training. In developing countries like Bangladesh, understanding and addressing the mental health needs of medical students is very important.

The purpose of this literature review is to critically examine existing studies pertaining to mental health, mood disorders, suicide prediction, and related factors among medical students. By analyzing current studies, this review aims to identify gaps, challenges, and opportunities for the development of a machine learning model to detect suicidal tendencies among the medical students in Bangladesh.

Several key themes and topics will be explored within the scope of this literature review. First and foremost is the prevalence of mental health issues among medical students, both globally and within the specific context of Bangladesh. By examining the prevalence rates and associated risk factors, an insight can be gained into the magnitude of the problem and the factors contributing to mental health challenges in this population.

Another major scope of this literature review revolves around medical students suffering from mood disorders, specifically depression and anxiety. Mood disorders are common among medical students and affect their academic success, personal and professional relationships, and general quality of life. Researching current information regarding these mood disorders, including prevalence, predisposing factors, and outcomes, can help to shed more light on the incidence and under-detection issue.

Subsequently, the literature review will investigate the scope of suicide prediction through examining extant models and approaches to identifying people at risk of committing suicide. Since suicide is a very tragic final stage of neglected mental disorders, it can affect anyone, including medical students. By studying recent strategies in predicting this suicide and their limitations, opportunities for creating an individually suitable and more efficient prediction model for medical students in Bangladesh can be recognized.

This literature review will pursue a multidimensional view of the factors that may impact the mental health outcomes of medical students. Primarily, such a multifactor approach would cover demographic conditions, economic conditions, family and social dynamics, health-related behaviors, academic experiences, and lifestyle conditions. The focus on the interaction of various elements would help to better understand the underlying issues. Through the integration of previous works, the research gap can be discovered. This approach will also help in guiding the workflow of this research while minimizing knowledge gaps. This effort is aimed at minimizing suicidal tendencies and remaining mental health determinants among this high-risk demographic.

2.1 Mental health

Mental health plays a significant role in an individual's well-being, and it is particularly relevant to medical students as they experience a variety of stressors and adversities during their academic path. This section investigates the mental health of medical students both within the global spectrum and Bangladesh by discussing the rate of individual mental health problems, major predisposing considerations, and the impact of untreated mental health conditions across different health outcomes.

The global prevalence rates of mental health issues among medical students are concerning, with high rates of depression, anxiety, and burnout observed in various countries. In Namibia, over one-third of medical students experienced depression or burnout [28]. Similarly, a mixed methods study conducted in Bangladesh revealed at least 50% of medical students suffer from moderate to severe stress, anxiety, and depressive symptoms [29]. Another study from the United States showed the increased risk of anxiety and depression among medical students [30]. Furthermore, this study also states that cultural, social, and environmental factors significantly influence mental health prevalence. While studies reveal that positive factors like study-life balance, meaningful relationships, and time with family positively impacts a student's well-being, the prior statement also means that less supportive educational climates, stigma towards mental health, and lower intentions to seek help, contribute to the prevalence of mental health issues among medical students.

Studies have shown the presence of many risk factors that contribute to mental health challenges among medical students. First, there are factors that belong to an individual level, including, but not limited to, academic stress, perfectionism, imposter syndrome related to academic performance [31]. Additionally, the second group of factors embraces institutional and systemic aspects like absence of mental health support services for young people, over competitive environment among medical students, high workload. In the third group, personal experiences, such as patient suffering and patient death, may contribute to mental health further [32]. The above-mentioned risk factors collectively lead to high levels of anxiety, depression, and burnout.

Depression, anxiety, and stress are common among students and have strong negative relationships with GPA and academic success [33]. Health anxiety was found to compromise grades, which indicates the importance of mental interventions to alleviate anxiety [34]. Mental health issues affect cognitive function, attention span, and motivation, which results in decline in the grades and scores for exams. Also, poor quality of sleep, anxiety, and depression result in decreased academic grades [35].

Students often apply adaptive coping strategies, including acceptance, positive reframing, and seeking support [36] which results in effective management of stress and mental health concerns. Maladaptive coping mechanisms including substance abuse and behavioral disengagement are less frequently utilized [37]. But coping styles may vary widely across individuals and depend on determinants such as gender, level of education, and overall health [38], influencing medical students' choices in dealing with stress. Hence, effective coping interventions and programs are needed for medical students. Specifically, solutions like crisis management training, peer support networks, and greater accessibility of mental health resources can effectively mitigate medical students' stress

and related issues [39].
Medical students have to face many barriers hindering their access to mental health support. Multiple studies have found stigma, fear of professional retaliation, confidentiality worries, and perceived ineffectiveness of services to be some important barriers [40, 41, 42]. The studies also further pointed out that cultural and institutional factors also play a role which need to be taken into account, and through destigmatization, enhancing accessibility, and integrating mental health services into medical education, a positive impact can be achieved.
Research in psychological science focuses on how gender, race, and other identities connect to shape mental health experiences [43]. These intersectional techniques demonstrate how gender, ethnicity, and systematic marginalization influence mental health in a variety of communities. Understanding these connections is critical for reducing inequities and developing effective solutions [44].
Untreated mental health issues among medical students can lead to long-term consequences. This, over time, may become chronic, with potential for relapse and progression to more severe conditions, impacting career satisfaction, professional competence, patient care quality and increasing risks of substance abuse, suicide, and academic attrition [45]. Therefore, early intervention is crucial to tackle these challenges and promote mental well-being of medical students.
Understanding the specifics of why medical students are more vulnerable to mental health issues is prominent. One study conducted a cross-sectional survey and the results when compared with postsecondary graduates from the general population, medical student respondents had significantly higher rates of psychological distress, suicidal ideation, and mood and anxiety disorders [46]. Another study found that medical students displayed worse psychological well-being and social relationships compared to the general sample [47]. Approximately half of the students reported low quality of life in the psychological and social domains, while a quarter experienced low quality of life in the physical health and environment domains. Although the phase of medical school did not have a significant impact on quality of life, gender differences were pronounced, with female students exhibiting worse physical and psychological well-being than their male counterparts.
There are also several evidence-based interventions and programs that have been formulated for promoting the mental wellbeing of medical students [48]: mindfulness-based interventions [49], and cognitive-behavioral therapy (CBT) programs [50] among others. Positive factors, such as study-life balance and academic achievement, supportive study relationships with peers or parents were reported to have a positive impact on students' psychological well-being. Conversely, the negative factor of educational stressors negatively impacts a student's psychological well-being. To integrate mental health promotion effectively, strategies like incorporating psychoeducation, CBT, mindfulness training, and peer support programs into medical curricula and institutional policies are crucial. Collaboration between academia, healthcare organizations, and community resources is essential to provide comprehensive support for medical students' mental health needs, as unaddressed mental health issues can lead to the development of mood disorders. The next section focuses on the different kinds of mood disorders that have a high impact on the lifestyle of medical students.

2.2 Mood disorder

Mood disorders are a subset of the overall mental health condition that is mainly associated with an individual's emotional state. It impacts a person's mood, interfering with their emotional well-being and daily functioning. Some common examples include depression, bipolar disorder, and suicidal ideation. Various studies around the world have highlighted high psychiatric morbidity and suicidal ideation, along with mood disorders, among medical students. In Bangladesh, the risk of medical students for depression and suicidal thoughts are significantly high; many studies reported its prevalence as 11% for suicidal ideation and 6.9% for attempted suicide in a Bangladeshi study, which is double than that of general population [51]. Bangladeshi university students, experiencing a high load of mental health problems and a significant rate of psychoactive medication use (14.4%), reported past-year suicidal ideation. Trends were also observed linking urbanicity and substance use as key risk factors during the COVID-19 pandemic [52]. As previously discussed, the significant risk factors that fuels mood disorders among medical students include academic stress, financial pressures, sleep deprivation, and exposure to sickness and death. These challenges, in turn, impact their academic performance, professional relationships, and personal well-being highlighting the critical need for comprehensive understanding and effective treatment across different populations along with discovering effective methods for the screening, detection and diagnosis of mood disorders.

One essential subset of mood disorders is depression, which must be thoroughly explored when discussing mood disorders. Notably, researchers have employed various methods to detect depression, including utilizing social media data [53, 54, 55, 56, 57, 58, 59], where Random Forest [53] and NLP techniques [54, 59] were employed. Additionally, studies explored unique approaches such as combining bi-LSTM with CNN [55, 56, 57], using questionnaires with CNN [60] and stacking algorithms [61], applying a hybrid of CNN and ANFIS to wearable sensor data [62], and using multiple algorithms to identify genes associated with depression [63]. The use of survey data for depression detection was found to involve techniques such as gradient-boosted decision trees [64], Random Forest [65], and the application of LSTM with CNN and NLP [66, 67].

Some studies leveraged the use of wearable sensors and mobile applications for assessing the level of depression [68, 69, 70] . Asare et al. used self-reported and wearable data with multiple AI algorithms for depression classification in one study [68] and in another study, utilized smartphone data and questionnaires with various AI algorithms for predicting depression [70]. On the other hand, Dogrucu et al. employed the Moodable app for distinguishing depression and suicidal ideation with diverse AI algorithms [69]. Furthermore, some studies have used questionnaires and survey data to assess the depression levels of participants [71, 72, 73, 74, 75, 76] . Lyall et al. employs different regression algorithms to identify sleep and circadian predictors [71]; Zhang et al. compares multiple algorithms for depression detection [72], and Siddiqua et al. used the combination of ten machine learning and two deep learning algorithms to predict depression reasons among students [73]. Two studies utilize Random Forest to assess depression levels, with one focusing on health-seeking behavior in major depressive disorder patients [74], and the latter on factors of depressive symptoms in Chinese elderly individuals [75]. Conversely, Uddin et al. employs LSTM and RNN to detect self-perceived symptoms in text [76]. Diagnosing depression is a challenge that has sparked the interest of inquisitive researchers, who employed unique techniques to try to diagnose depression among 9 individuals using unique AI models. In some studies, EEG datasets were utilized [77, 78, 79]. One study utilized linear discriminant analysis (LDA) and linear support vector machine (SVM) to predict depression in drug-naive patients [77], one employed multiple algorithms [79], while another study introduced a hybrid model with bi-LSTM for depression diagnosis [78]. Furthermore, researchers have also leveraged data extracted from wearable devices and mobile applications for diagnostic purposes [80, 81, 82]. They individually utilized different approaches like DCNN [80], a combination of multiple algorithms [81] and neural networks with weighted fuzzy membership functions (NEWFM) [82] on data extracted from wearable sensors to make accurate diagnostic decisions.

Screening for depression is a critical process aimed at early identification and intervention for individuals. Thus, studies have made use of various types of data to be precise with its identification, by incorporating AI models on them [83, 84, 85, 86, 87]. Wani et al. focused on using deep learning to identify depressive symptoms in netizens through social media content [83], while Geng et al. optimized initial Major Depressive Disorder (MDD) screening using 5-minute ECG signals during sleep through multiple algorithms [84]. Another study aimed to detect depression risk in a working population with digital biomarkers from wearables [85]. Additionally, [86] predicted depression scores in adolescents using a mix of linear and non-linear machine learning techniques. In contrast, [87] enhanced MDD screening using multiple algorithms in the EarlyDetect mobile app.
The treatment of depression is usually a combination of different techniques due to its complex nature. Accordingly, studies have shown promising results in tracking, monitoring, and providing optimized treatment [88, 89, 90]. One study utilizes multiple algorithms to predict MDD using app, sensor, and questionnaire data [88], while another study observes ketamine effects on MDD patients by applying multiple AI algorithms on genome-wide sequencing data [89]. Additionally, Mullick et al. applied LSTM and a fully connected network to predict optimal medication for individuals with diagnosed depression [90].
In assessing depression severity with Bengali text data, one study uses a mix of machine learning and deep learning approaches, showcasing a diverse strategy [91]. Alternatively, another study develops the PSYCHE-D model employing multiple algorithms on person-generated health data for forecasting long-term depression severity increases [92]. Additionally, a separate investigation utilizes a Gated Recurrent Unit (GRU) within a deep multitask framework, employing machine learning models to detect depression, sentiment, and multi-label emotion from suicide notes [93].
Another very common type of mood disorder is bipolar disorder characterized by manic, hypomanic, or mixed episodes alongside depressive episodes, affecting approximately 0.4% to 1% of the global population [94]. Patients with bipolar disorder often experience significant functional impairment and medical students are no exception. Research indicates that Bipolar Disorder is relatively common among medical students, with prevalence rates ranging from 4.5% to 26.84% [95]. Although this condition can be managed effectively with appropriate treatment, the fear of stigma, attitudinal barriers, and systemic obstacles often hinder medical students from seeking help for mental health problems, potentially leading to self-harm and suicide [96]. Accordingly, researchers have experimented with many methodologies to try and tackle this issue through novel approaches. For early detection of bipolar disorder using questionnaire data, Bennett et al. compared the performance of multiple machine learning algorithms to observe PHQ score fluctuations [97], and Liu et al. used elastic nets, a regression model using EarlyDetect questionnaire data [98]. Busk et al. used Bayesian regression on self-assessment data to predict mood scores of patients suffering from bipolar disorder and a support system to monitor patient condition and avoid hospitalization [99]. Another supporting model for patients suffering from bipolar disorder used multiple machine learning algorithms on the last 6 hospital visit data of individuals to predict their depressive relapse [100].
When mood disorders like depression and bipolar disorder reach a certain severity without proper treatment, patients may harm themselves, and it can often lead to suicide. In the next section, research surrounding suicidal tendencies will be discussed in detail.

2.3 Suicide tendency

Suicidal tendencies are characterized as a range of actions, from suicidal thoughts to suicide. Suicidal ideation refers to thoughts of self-harm or suicide, whereas a suicide attempt is a determined effort to inflict self-harm with varied degrees of severity. Completed suicide is the terrible end in which the individual dies as a result of self-inflicted wounds. Global data on suicidal inclinations show considerable differences across populations and geographies. According to studies, almost 800,000 individuals die by suicide each year, with low- and middle-income nations accounting for roughly 74% [101]. However, implementing these characteristics takes time and, on their own, may not be sufficient to significantly reduce the number of suicides. And this has not deterred academics from seeking the best answer for this problem.
For detecting suicidal ideation, deep learning Recurrent Neural Network (RNN) algorithms such as LSTM, BERT, and GRU along with CNN were used on social media text data [102, 103, 104] and health records containing medical and clinical data [105]. Alternatively, the performance of multiple 10 supervised machine learning algorithms was compared using social media data [106, 107] and survey data [108] for the detection of suicidal ideation among people.
For the assessment of patients, the performance of multiple machine learning classification and regression algorithms was compared using health records data [109, 110, 111], survey data [112, 113], and social media data [114, 115] to develop screening tools for suicidal ideation and self-harm behavior. Lu et al. developed a transformer encoder patient assessment model using unstructured data to improve the prediction of suicide and self-harm [116], whereas Cho et al. developed a similar model for suicide risk using random forest [117]. Edgcomb et al. used CART to develop a hospital admission screening tool that predicts the risk of hospital readmission due to suicidal behavior [118]. To distinguish different mental states in suicidal individuals, Wang et al. developed a long-term and short-term suicide risk prediction model using a Light gradient boosting machine [119] and Su et al. developed a patient assessment tool that can distinguish between self-harm tendency and suicidal ideation to identify unique risk profiles [120]. Suicide risk screening models were also developed using speech data by analyzing voice quality, features, and fluency by comparing and evaluating the performance of multiple machine learning algorithms [121, 122]. To provide support and prevent suicide in individuals, Hechler et al. developed a mobile application called Thoth that utilizes wearable sensor data, social media data, and questionnaire information to predict the risk of suicidal ideation by applying and comparing multiple machine learning algorithms where GRU showed the best performance [123].

2.4 Conclusion

In conclusion, the study on the mental health situation of medical students worldwide and in Bangladesh depicted an alarming scenario. The frequency of different mental health conditions has been very high, as well as the relative and absolute risk factors. Inequality in class, gender differences, individual and institutional factors proved to be the important causative factors. Studies clearly implicated that the mental health conditions of medical students tend to progress and become chronic if not treated in time. These untreated mental health problems affect not only academic performance but also professional competence, career satisfaction, and overall well-being. Any efforts to support the mental health of medical students must incorporate policies, information, interventions, and approaches that are open-ended. These include reducing stigma, increasing awareness, creating clear paths to mental health support services, and targeting supportive interventions to support individuals cope with stress in adaptive ways. Short term, non-enduring interventions, like mindfulness-based interventions and cognitive-behavioral therapy (CBT), have a valuable niche and essential role in the overall management and wellbeing of medical students. At an advanced stage, intersectional research is needed to gain a better understanding of the unique experiences of medical students in terms of gender, ethnicity, and other identities. This can inform future interventions.

Mood disorders, such as depression and bipolar disorder, cause significant challenges to an individual's mental health, and medical students are no exception. Detecting and diagnosing these disorders require innovative approaches, where the use of artificial intelligence and wearable technology can be beneficial. Early intervention and comprehensive treatment are essential for managing mood disorders effectively and preventing adverse outcomes like self-harm and suicide. Table 1 presents the prevalence of suicide worldwide, based on a study conducted in 2022 [124].

Table 1: Suicide Rates in Major Countries (per 100,000 population)

Country	**Suicides per 100,000**	**Country**	**Suicides per 100,000**
Lesotho	87.5	Guyana	40.9
Eswatini	40.5	Kiribati	30.6
Micronesia	29	Suriname	25.9
Zimbabwe	23.6	South Africa	23.5
Mozambique	23.2	Russian Federation	21.6
Republic of Korea	21.2	Vanuatu	21
Botswana	20.2	Lithuania	20.2
Uruguay	18.8	Kazakhstan	18.1
Mongolia	18	Ukraine	17.7
Solomon Islands	17.4	Eritrea	17.3
Belarus	16.5	Montenegro	16.2
Latvia	16.1	Cameroon	15.9
Côte d'Ivoire	15.7	Cabo Verde	15.2
Togo	14.8	Somalia	14.7
Samoa	14.6	USA	14.5
Burkina Faso	14.4	Zambia	14.4
Slovenia	14	Belgium	13.9
Equatorial Guinea	13.5	Namibia	13.5
Finland	13.4	Chad	13.2
Gabon	13.1	India	12.9

Suicidal tendencies have been found to grow among medical students according to trends, which necessitates proactive measures for detection, assessment, and prevention. Utilizing machine learning algorithms on various data sources can help to aid in identifying and providing timely support. The most effective machine learning models addressing suicide, as identified in the referenced studies, are compared in Table 2.

Table 2: Comparison of machine learning models used in previous similar researches.

Reference	**Algorithms Used**	**AUC**
105	Convolutional Neural Network	0.988
108	Logistic Regression	0.83
110	Random Forest	0.7
111	Convolutional Neural Network	0.946
112	Logistic Regression	0.832
113	Logistic Regression	0.77
114	Support Vector Machine	0.68
115	Random Forest, Support Vector Machine	0.86, 0.86
117	Random Forest	0.81
118	Decision Tree	0.86
119	Gradient Boosting	0.91
120	Random Forest	0.89
121	Artificial Neural Network	0.61
129	Ensemble	0.9

Thus, a predictive model is brought forward here that further enhances early detection and intervention strategies, ensuring that individuals at risk receive the appropriate mental health care and support promptly.

Chapter 3

Exploration of predicting suicide ideation

This chapter explores the various techniques adopted for the prediction of suicide ideation. It focuses on the individual, relational, societal, and cultural factors that affect risk. Moreover, it addresses the specific challenges faced by individuals and the role of academic pressure, family dynamics, and mental health and its influence on suicidal thoughts. Here, the diverse data collection methods are highlighted such as self-report questionnaires, clinical interviews, and medical records, along with the use of standardized tools like the PHQ-9 and GHQ-28 for assessing suicide risk. Furthermore, to predict suicide ideation, application of various machine learning algorithms have been explored that uses social media, clinical, and unstructured data. Key evaluation metrics such as accuracy, sensitivity, specificity, and area under the ROC curve are also talked about in this chapter. Finally, the need to integrate cultural considerations and advanced data analytics into suicide prevention strategies are mentioned in order to develop effective, data-driven interventions.

3.1 Suicide Ideation Risk Factor Analysis

Suicide risk is influenced by a combination of individual, relational, community, and societal factors. Some key risk factors identified in the literature include academic strain, family dynamics, socioeconomic position, traumatic experiences, substance abuse, and psychological distress. Individuals face unique challenges that act as a catalyst to these general risk factors. According to three separate studies, intense academic pressure and demanding parents are significant risk factors, leading to suicidal ideation among students [52, 125, 126]. However, according to a contradictory meta-analysis, negative mental health outcomes such as feeling depressed, burnout, having other mental health problems at the same time, and experiencing stress were found to be the symptoms associated with suicidal ideation and suicidal attempts. In contrast, factors such as smoking, a family history of mental illness, and past suicidal behavior did not show a substantial increase in the risk of suicidal thoughts [127].

3.2 Data collection and Screening tools

Collecting such data requires accurate data collection strategies. Common methods include surveys and self-report questionnaires, clinical interviews, academic performance records, peer and faculty observations, and electronic health records. Furthermore, several standardized scales are also being used in suicide risk research and as screening tools. One study estimating the prevalence of depression, depressive symptoms, and suicidal ideation among students employed the Patient Health Questionnaire (PHQ-9) as well as the General Health Questionnaire (GHQ-28) [128]. For the purpose of screening suicidal ideation in a prospective cohort study among medical interns, assessment tools such as PHQ-9, Risky Families Questionnaire, Generalized Anxiety Disorder 7 (GAD-7), Medical Outcomes Study Short-Form Health Survey (SF-8), Epworth Sleepiness Scale (ESS), and the Perceived Stress Scale (PSS) were used. Figure 1 shows the data collection tools used in studies on suicidal ideation.

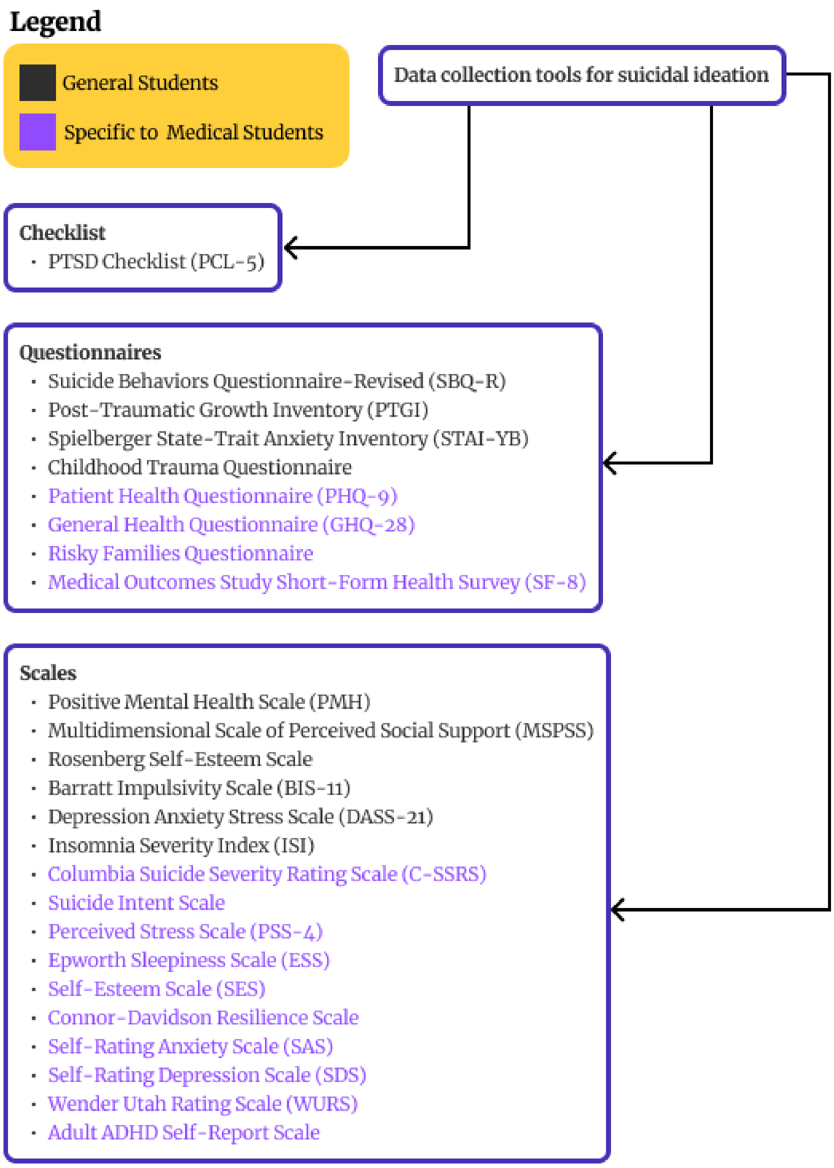

Figure 1: Data collection tools used in studies predicting suicidal ideation.

3.3 Cultural Specific Data Considerations

Understanding cultural context is crucial when analyzing suicide risk data. Different cultures have varying stigmas and perceptions of mental health, which can affect how individuals report symptoms and seek help. For example, in some cultures, mental health issues are highly stigmatized, leading to underreporting and reluctance to seek treatment. In Bangladesh, societal pressures and expectations, particularly regarding academic and professional success, can contribute significantly to mental health issues [125]. Furthermore, parental divorce, extra marital issues, blackmail, and career restricting disabilities are unique suicidal factors of Bangladesh. In Canada, factors such as cigarette use, clerkship, family history of mental illness, family history of suicide attempt, living away from home, and ragged by others remain unique [127], while the year of study, age and negative emotions are unique to China [129]. Accordingly, every country possesses unique factors specific to its context, and this must be considered beforehand to make sure interventions are culturally relevant, sustainable, and effective.

3.4 Types of data used in Machine Learning models

This section discusses the data needed to develop, implement, and test intervention models targeting suicide ideation. Studies in Section 2.4 highlight various data types used in creating suicide-related machine learning models, including social media, questionnaires, medical records, and more complex unstructured data like images and voice recordings. Figure 2 illustrates the frequency of those data types in machine learning models found in the literature.

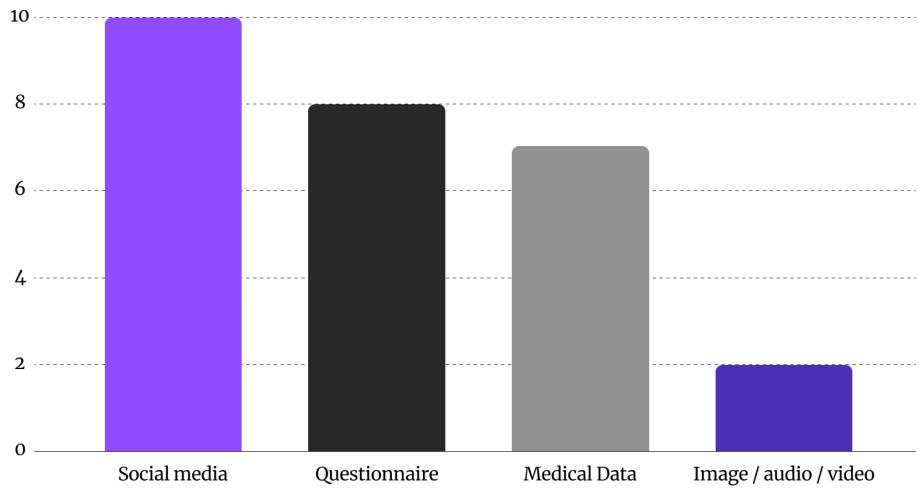

Figure 2: Data collection mediums used in predicting suicidal ideation.

In the context of medical students, data needs to be carefully curated to identify factors contributing to mental health challenges and suicidal ideation in an academic, medical, and social setup. Accordingly, one study implemented a stacked ensemble Decision Tree on demographic data extracted using questionnaires (PHQ-9, SBQ-R, PTGI, PMH, MSPSS) and a checklist (PCL-5) for accurate diagnosis of suicidal ideation [130]. In order to identify the medical lethality of suicide attempt, another study implemented Random Forest, Linear Regression, and Convolutional Neural Network (CNN) on the Columbia-suicide severity rating scale (C-SSRS), Suicide Intent Scale (SIS) and collected demographic and clinical data [131]. Screening of suicidal behaviors of university students in Bangladesh regarding COVID-19 were assessed using Logistic regression (LR), Support Vector Machine (SVM), Naïve Bayes (NB), Classification Tree (CT), and Random Forest (RF) on demographic, socioeconomic and academic data collected through SBQ-R, DASS-21, ISI [132]. Moreover, a study on predicting the probability of suicide attempts among Chinese medical college students applied Random Forest on demographical and clinical data extracted through utilizing SAS, SDS, ESS, WURS, SES, Adult ADHD Self-Report Scale, and the Connor-Davidson Resilience Scale [133]. These scales extract various demographic, educational, economical, general health related, stress-related, mental health related data and more from populations, most of whom are located in High Income Countries (HIC).

3.5 Types of algorithms used in suicide prediction

In the studies discussed in the literature, it was found that Random Forest was used the most to predict suicidal ideation. The second most prevalent was SVM where it utilized social media data in 4 studies, medical data in 2 studies, speech data in 2 studies, and questionnaire data in 1 study. Other forms of supervised algorithms like logistic regression, Gradient Boosting, XGBoost, Naive Bayes and Bayesian networks were found to be used in some studies working with medical data or questionnaires. On the other hand, unsupervised algorithms such as CNN, LSTM, GRU, NN and ANN sought determinants of suicidal ideation through social media data. The algorithms used in predicting suicidal tendency is shown in Table 3.

Table 3: Studies using different types of machine learning algorithms to train suicidal tendency prediction models.

Reference	**RF**	**SVM**	**LR**	**DT**	**NB**	**BN**	**GB**	**RNN**	**CNN**	**AE**
[102, 103, 105]								✓	✓	
[104, 123]								✓		
[106]	✓	✓	✓						✓	
[107]	✓	✓					✓			
[108]			✓	✓	✓					
[109]	✓		✓							
[110]	✓	✓				✓		✓	✓	
[111]		✓	✓		✓				✓	
[112]	✓	✓	✓				✓			
[113]	✓					✓	✓			
[114, 121]	✓	✓					✓			
[115]	✓	✓								
[116]										✓
[117, 120]	✓									
[118]				✓						
[119]							✓			
[122]		✓								

3.6 Types of evaluation used in suicide prediction

While a lot of evaluation metrics were used, the most common were accuracy, precision, sensitivity, specificity, F1 score and area under the ROC curve as seen in Table 4.
The proportion of correctly classified instances out of the total, which provides an overall indication of the model's correctness, is measured by accuracy:

$$\text{Accuracy} = \frac{TP + FP}{TP + FP + TN + FN} \tag{3.1}$$

(how many instances was the model correct about)

The true positive rate, represented by sensitivity or recall, shows how well the models identified instances in the positive class and is crucial for tasks like identifying suicidal thoughts:

$$\text{Sensitivity / Recall} = \frac{TP}{TP + FN} \tag{3.2}$$

(how many instances were correctly identified in positive class)

Specificity, the true negative rate, indicates a model's ability to correctly identify instances in the negative class, providing a sense of the model's ability to avoid false alarms:

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3.3}$$

(how many instances were correctly identified in negative class)

Precision, measuring the accuracy of positive class predictions, highlights how well the models performed when predicting instances as positive:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.4}$$

(how well did the model perform in it prediction of positive class)

F1 score, representing the harmonic mean of precision and recall, offers a balanced measure of a model's performance, especially in scenarios with imbalanced class distributions:

$$\text{F1 Score} = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3.5}$$

(how well did the model perform in it prediction of positive class)

The ROC curve graphically illustrates the trade-off between sensitivity and specificity, allowing a visual understanding of each model's discrimination threshold. The AUC, representing the area under the ROC curve, provides a single metric to compare model performance, with higher AUC values indicating better overall performance. Figure 3 demonstrates the Area under the ROC curve.

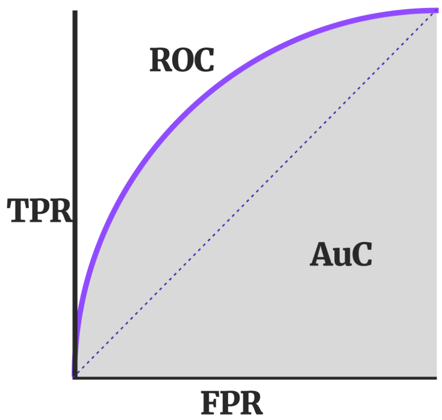

Figure 3: Depiction of the AuC-ROC. (TPR = True positive rate, FPR = False positive rate, ROC = Receiver-operating characteristic, AuC = Area under curve)

Table 4: Studies using different metrics used to evaluate suicidal tendency prediction models.

Reference	Accuracy	Recall	Precision	Specificity	F1 score	AUC-ROC
[102], [105], [106], [107], [111], [114], [115]	✓	✓	✓		✓	✓
[116], [117], [119], [120]		✓		✓		✓
[110], [118], [123]	✓	✓		✓		✓
[103], [107]	✓	✓	✓		✓	
[104], [109]	✓	✓		✓		
[121]	✓	✓	✓	✓	✓	✓
[113]		✓				✓
[108]		✓		✓		✓
[122]	✓				✓	
[123]	✓					
[112]						✓

3.7 Conclusion

In responding to this concerning phenomenon of suicidal ideation, particular attention should be paid to the composition of the overall risk factors associated with this group and the specific difficulties faced by learners. It is imperative that with the data gathered, practical and advanced screening tools should be employed in addition to consideration of some cultural factors that could be relevant in the development of these models. Random Forest and SVM conceptual models have revealed potential in predicting at-risk persons with several sorts of data, including social media data, medical history, and self-report data. By the application of these progressive approaches and technologies, it is possible to promote a better comprehension of the subject of suicidal ideation and apply deeper interventions designed for boosting the psychological wellbeing worldwide.

Chapter 4

Methodology

This chapter outlines the comprehensive approach to predicting suicide ideation risk among medical students in Bangladesh. First, the data collection process has been detailed, which yielded three distinct datasets from Bangladesh and Thailand containing various mental health indicators and suicidal attributes. The chapter then explores the systematic data preprocessing techniques adopted, including handling missing values, encoding categorical variables, and normalizing numerical features. Next, the employment of four feature selection methods: LASSO regularization, Pearson's correlation, Sequential Forward Selection, and Weighted Sum Feature Selection are shown to identify the most relevant predictors. The chapter further elaborates the implementation of four supervised machine learning algorithms (Logistic Regression, Random Forest, Support Vector Machines, and Decision Trees), discussing the specifics of hyperparameter tuning for optimal performance. Finally, the evaluation metrics are presented, which includes accuracy, sensitivity, specificity, precision, F1 score, AUC, and Cohen's Kappa to comprehensively assess model effectiveness.

4.1 Data Collection

Medical student datasets from low and middle-income countries (LMIC) were searched for thoroughly, and two datasets from Bangladesh and one from Thailand were acquired successfully. These datasets met our criteria for including medical data with suicidal attributes, either through questionnaires or calculated scales. Initially, the search was conducted for datasets of Bangladesh or LMIC origin that met our criteria. However, due to class imbalance, search was continued for additional datasets but eventually halted when no more public datasets were available.

The first dataset, Dataset A, conducted a cross-sectional study on mental health, suicidal behaviors, and suicide literacy among Bangladeshi undergraduate dental students. Here, a total of 487 students participated, with 468 included in the final analysis, utilizing a questionnaire covering various aspects such as socio-demographics, psychiatric history, depression, anxiety, suicidal behaviors, literacy, and stigma [134]. This dataset had no symptom based item included, instead it focused on collecting various categories of data such as Demography, Lifestyle, Economical, Educational, Medical History, Mental Health History and Family History data. Total attributes in Dataset A are 37 among which "Suicidal Plan" has been selected as the target variable. Table 5 shows the first 4 items of the dataset with all the features.

Table 5: Features and data of first 4 rows from Dataset 1

Features	**Row 1**	**Row 2**	**Row 3**	**Row 4**
Gender	Female	Female	Male	Male
Location	Outside	Outside	Inside	Inside
Year of study	Third Year	First year	Third Year	Second Year
Permanent residency	Urban	Urban	Rural	Rural
Religion	Islam	Islam	Islam	Islam
Family type	Nuclear	Join	Nuclear	Join
Family monthly income	More than 30000	More than 30000	Less Than 15000	More than 30000
Cigarette smoker	Not Smoker	Not Smoker	Not Smoker	Not Smoker
Psychoactive substance user	Not	Not	Not	Not
Chronic physical disease since last year	No	No	No	No
Family mental illness history	Yes	No	No	No
Family suicide history	No	No	No	No
Family suicide attempt history	No	No	No	No
Chronic psychological sufferings since last year	Yes	No	No	No
Post-traumatic events since last year	Yes	No	No	No
Suicidal Attempt Lifetime	No	No	No	No
Suicidal Plan Month	No	No	No	No

The second dataset was curated from November 2020 to December 2020 at Enam Medical College and Hospital which is situated in Savar, Dhaka, Bangladesh. It is a web-based questionnaire to collect data from private medical students, employing the Beck Depression Inventory II (BDI-II) scale to measure depression and suicidal ideation to perform a cross-sectional web based survey [135]. The dataset collected many symptom based data alongside various categories of data such as Demographic, Lifestyle, Economical, Educational and Family History. Total attributes in Dataset B are 35 among which “Suicidal Thoughts” has been selected as the target variable. This is the 9th item from the Beck Depression Inventory II (BDI-II). Table 6 shows the first 4 items of the dataset with all the features.

Table 6: Features and data of first 4 rows from Dataset 2

Features	**Row 1**	**Row 2**	**Row 3**	**Row 4**
Gender	Male	Female	Male	Female
Age	20–23 years	24 years and above	20–23 years	20-23 years
Marital Status	Single	Married	Single	Single
Family Type	Adjunct Family	Adjunct Family	Adjunct Family	Adjunct Family
Family Problems?	Absent	Present	Present	Absent
Family History of Depression	Yes	Yes	No	Yes
Additional Earnings	No	Yes	No	No
Study Year	1st	5th	3rd	1st
Failed Subjects	No	No	Yes	No
Monthly Family Income	Above 120000 Tk	81000-120000 Tk	10000-40000 Tk	81000-120000 Tk
Alcohol Use	None	None	None	None
Drug Addiction	None	None	None	None
Relationship Status	In a good relation	Never had any	Never had any	Never had any
+ 21-items of the Beck Depression Inventory II (BDI-II)				

The third dataset, Dataset C, was created as part of a cross-sectional study to investigate the prevalence and associated factors of depression in medical students from Northern Thailand, Chiang Mai University [136]. The authors of this dataset extracted general demographic data such as sex, age, year of study, hometown, relationship status, financial status, accommodations, underlying diseases (including mental health difficulties and substance usage), and family psychiatric history. Furthermore, they also extracted psychiatric problems, depression and protective factors using multiple scales. Total attributes in Dataset C are 152 among which “Suicidal Ideation” has been selected as the target variable. Table 7 shows the first 4 items of the dataset with all the features.

Table 7: Student Characteristics and Survey Responses

Features	**Row 1**	**Row 2**	**Row 3**	**Row 4**
Gender	female	male	female	female
Age	17	18	19	19
Hometown	Bangkok	Chiang Mai	Chiang Mai	Northern
Repeater	no	no	no	no
Educational Year	1	1	1	1
Relationship Status	Other	Other	Other	Single
Income	5000	7000	7000	5000
Enough Income	yes	yes	yes	yes
Live With	parent	Alone	Alone	parent
Living Place	House	House	House	Apartment
Psychiatric Consultation	no	no	no	no
Psychiatric Disorder	no	no	no	no
Mood	no	no	no	no
Other Psychiatric Disorder	no	no	no	no
Medical Comorbid	no	no	no	yes
Major Depressive Disorder	no	no	no	no
Bipolar	no	no	no	no
GAD	no	no	no	no
ADHD	no	no	no	no
Adjustment	no	no	no	no
Other (Eating / Trichotillomania / OCD / Insomnia)	no	no	no	no
Asthma	no	no	no	no
Allergy	no	no	no	yes
Dyspepsia	no	no	no	no
Thyroid	no	no	no	no
PCOS	no	no	no	no
G6PD	no	no	no	no
Other Medical Comorbid (HT/DM)	no	no	no	no
Drink Frequency	no	yes	yes	Once a week
Drink Amount	no	no	no	yes
Peer Pressure Alcohol	no	no	no	yes
Stress Alcohol	no	no	no	no
Insomnia Alcohol	no	no	no	no
Ever Smoke	Never	Never	Never	Never
Reason Smoking	no smoking	no smoking	no smoking	no smoking
Peer Pressure Smoke	no	no	no	no

(Continued on next page)

(Continued from previous page)

Features	Row 1	Row 2	Row 3	Row 4
Stress Smoke	no	no	no	no
Other Reason Smoke	no	no	no	no
Ever Use a Substance	no	no	no	no
Family Psychiatric Problem	no	no	no	no
Current Psychiatric Symptoms	no	no	no	no
Treating Psychiatric Symptoms	no	no	no	no
Onset of Psychiatric Symptoms	no psychiatric problem	no psychiatric problem	no psychiatric problem	no psychiatric problem
Stress Due to Education	no	yes	yes	yes
Stress Due to Finance	no	no	no	no
Stress Due to Friend Relationships	no	yes	yes	yes
Stress Due to Romantic Relationships	no	no	no	no
Stress Due to Family Relationships	no	no	no	yes
Stress Due to Teacher Relationships	no	no	no	no
Stress Due to Physical Health	yes	no	no	yes
Stress Due to Mental Health	yes	no	yes	yes
Stress Due to Living Place	no	no	no	no
Stress Due to Future/Career	no	no	no	no
Stress Due to Other	no	no	no	no
Stress Due to Relationships	yes	yes	yes	yes
Stress Due to Health	yes	no	yes	yes
Stress Due to Others	no	no	no	no
Difficult Social Relations	4	7	4	yes
Enough Grade	8	10	9	6
Bore to Learn	7	4	4	7
Apply to Wrong Faculty	no	no	no	no
Emotional Exhaustion	high	high	low	low
Depersonalization	high	high	high	low
Personal Efficacy	low	high	high	high
Burnout	no	no	no	no
Current Game	yes	yes	no	no
Suicidal Ideation	Several days	Not at all	Not at all	Several days

Some important attributes of the three datasets are shown in Table 8. Furthermore, Table 9 contains the categories of data collected by each dataset.

4.2 Data Preprocessing

All the datasets went through the same data preprocessing techniques to ensure uniformity. This process included handling missing values, analyzing the collected data, removing duplicates & redundant data (scales and questionnaire), label encoding and normalization. In this section, the

Table 8: Attributes of the selected datasets.

Datasets	**Bangladesh Data**	**Medical Student Data**	**Publication Backed**	**Data Columns**	**Total Items**	**Positive Class**	**Negative Class**
Dataset A	✓	✓	X	37	468	15	453
Dataset B	✓	✓	✓	35	237	172	65
Dataset C	X	✓	✓	152	706	88	618

Table 9: Categories of data covered by the selected datasets.

Categories	**Dataset A**	**Dataset B**	**Dataset C**
Demographic	✓	✓	✓
Lifestyle	✓	✓	✓
Economical	✓	✓	✓
Educational	✓	✓	✓
Symptom	X	✓	✓
Medical History	✓	X	✓
Mental Health History	✓	X	✓
Family History	✓	✓	X

preprocessing of the collected data will be explained in detail.

All datasets were carefully inspected for missing values manually, as the absence of information can introduce potential bias. Dataset A had 3 rows with missing values, which were subsequently dropped because the values represented a small percentage of the overall dataset and their removal would have minimal impact on the analysis. On the other hand, Dataset B and Dataset C had no missing values. For the next step, all the features were plotted as a bar graph individually for each dataset to visualize the distribution of the classes. Valuable insights such as the "equal number of male and female participants", "low overall count of alcohol intake" among participants were identified from all the datasets. Such information about the attributes or columns helped to make decisions in the upcoming steps. Furthermore, this step was crucial in identifying redundant values from the dataset that were identified and removed.

Since the majority of data items in the datasets were categorical, label encoding was incorporated. This method was selected based on the previous data analysis step to ensure consistency and control over the encoding process for each column.

Finally, after incorporating label encoding for the categorical data, Min-Max normalization was performed on the numerical features. Min-Max normalization scales the data to a specific range (typically 0 to 1), preserving the relationships between the values while putting them on a common scale. This normalization was applied to ensure that all features contribute equally to the analysis and prevent any feature from dominating the model due to its scale.

4.3 Feature Selection

Feature selection is crucial in machine learning as it helps in improving model performance by selecting the most relevant features while reducing noise and redundancy [137]. In this study, four feature selection techniques namely LASSO or L1 Regularization, Pearson's Correlation Coefficient, Sequential Forward selection, and Weighted Sum Feature Selection were considered. Below are the detailed descriptions and comparisons of each method. The Least Absolute Shrinkage and Selection Operator (LASSO), which is an embedded method, attempts to select features that have a significant impact on the prediction. It performs feature selection and regularization simultaneously by introducing a penalty term which directs the coefficients of less important features to shrink to zero. The Pearson's correlation coefficient has been implemented as the filter method which selects features based on the correlation between each variable and the target variable. The features with the highest correlation coefficient are considered for further steps. This method helps to identify the features that have a notable impact on the target variable from a statistical point of view. The sequential feature selection technique (SFS), a wrapper method, has been selected due to its efficiency in selecting relevant features while maintaining good performance [138]. This technique incrementally adds features to an empty set until a stopping criterion has been reached. This ensures that the model is concise yet effective, improving interpretability and computational performance [139]. However, Sequential Forward Selection (SFS) is a greedy approach that generates locally optimal choices rather than a globally optimal solution. It selects the locally optimal feature set without considering the global effect. As a result, it often fails to consistently identify the most relevant features as the training sets change, and it is frequently unable to eliminate redundant features from the feature space.

Thus, a Weighted Sum feature selection algorithm has been chosen as the final feature selection strategy for comparison [140]. WSFS overcomes the limitations of SFS by selecting the most relevant feature independently with changes captured in the training set. This approach involves partitioning of the training set into k folds for cross validation and assesses the current fold using SFS with ranked features for identifying suboptimal features set. When each of the evaluations is complete, the features involved are weighted according to the contribution made, and a total weight is then derived. Lastly, all the features are divided into k overlapping partitions with the help of total weight and then the subset with maximum accuracy is selected as the optimal feature set of classification.

4.4 Model Training

Before moving towards feature selection, it is important to make sure that our target variable is balanced, which is crucial for mitigating bias and ensuring fairness in decision-making processes. Studies emphasize that imbalance in training data, particularly in protected attributes, can result in higher levels of unfairness in the output [141]. The Synthetic Minority Oversampling Technique (SMOTE) has been implemented to overcome the issue of oversampling and learning from imbalanced data as it is considered to be one of the most influential data preprocessing and sampling algorithms in machine learning and data mining [142]. Finally, after balancing the positive and negative classes of the target variable, the model is then ready to be trained.

Four supervised learning algorithms were selected, namely Logistic Regression (LR), Random Forest (RF), Support Vector Machines (SVM) and Decision Trees (DT) to train the model.

For the environment setup, the Jupyter Notebook was utilized along with the help of the scikit-learn library.

4.4.1 Logistic Regression

LR is a simplified version of the Generalized Linear Model (GLM) operator, which uses probability ideas to connect an independent variable (x) to a dependent variable (y). In binary regression, the input values (x) usually indicate two classes (0 or 1) [143]. The input values are merged linearly with coefficient values or weights to get an output value (y). This method creates a regression model to estimate the likelihood that input data belongs to a given labeled class, which might be '0' or '1'. One such study involving a questionnaire dataset of Korean adolescents applied multiple ML algorithms, where LR delivered the best performance [144]. Through the use of the sigmoid function of logistic regression, the study predicted the odds ratio for each observation to be classified as positive or negative.
In this study, an L1 penalty was utilized for regularization, known as Lasso regularization. L1 regularization adds a penalty equal to the absolute value of the coefficients' magnitudes, encouraging simpler models with fewer non-zero coefficients. Additionally, the 'liblinear' solver was employed, chosen for its effectiveness with small datasets and its compatibility with L1 regularization, making it ideal for our dataset. Lastly, the 'max_iter=1000' parameter was set to control the maximum number of iterations for the solver to converge in the logistic regression algorithm, ensuring convergence within a specified limit.

4.4.2 Random Forest

The random forest algorithm (RF) based on the aggregation of ensembles of decision trees created through recursive bootstrapping [145]. They are a popular ensemble method in machine learning, known for their ability to handle complex data effectively. They have been extensively utilized in various domains such as medicine, microbiology, and symptom diagnosis [146]. Accordingly, one study conducted a survey among medical students and implemented Random Forest on the collected data to effectively and accurately identify the probability of suicide attempts [133].
In this study, the Random Forest classifier is defined without any specific hyperparameters, allowing them to be tuned later. GridSearchCV is then utilized with a predefined grid of hyperparameters (n_estimators, max_depth, min_samples_split, and min_samples_leaf) to perform an exhaustive search for the best combination of these parameters through cross-validation. The 'n estimators' parameter determines the number of trees in the forest, 'max depth' controls the maximum depth of the trees, 'min_samples_split' sets the minimum number of samples required to split an internal node, and 'min_samples_leaf' specifies the minimum number of samples needed to be at a leaf node. This grid of hyperparameters is used in conjunction with GridSearchCV, which exhaustively searches through these combinations via 5-fold cross-validation to find the optimal set of hyperparameters that yields the best model performance.

4.4.3 Support Vector Machine

Support Vector Machines (SVM) is a supervised machine learning technique designed for binary classification tasks, where it works by separating data points in a high-dimensional space into two classes using a hyperplane. This hyperplane is positioned to maximize the gap, known as the 'margin,' between the hyperplane and the support vectors. The traditional SVM is effective for linearly separable data [147]. Additionally, the kernel coefficient (gamma) influences the shape of the decision boundary, impacting the model's flexibility and generalization capability. Many studies have made use of SVM's capability of handling both classification and regression tasks.

One study even concluded that SVM can be used to develop the best predictive algorithm to identify at-risk university students with suicidal ideation/behaviors [132].
For this study, all datasets were preprocessed by standardizing the features using StandardScaler. A linear kernel was used and to control the regularization strength, GridSearchCV was used to tune the SVM's hyperparameter 'C'.

4.4.4 Decision Trees

The Decision Tree (DT) model uses a tree-like structure to analyze data and determine target values [148]. It segments data recursively until all points are in specific classes. The tree consists of a root, internal nodes with test conditions, and leaf nodes with class labels. Node creation follows a top-down approach with two stages: tree construction and pruning [149]. Predictions are based on majority samples at leaf nodes during training, and values in leaves are averaged for numerical predictions. One study used a type of decision tree to assess predictor combinations, automate feature selection, and construct a model optimized for clinical interpretability in a computationally efficient way [150].
In this study, the features of the datasets were standardized first using StandardScaler, The script then proceeds to use GridSearchCV to search through a grid of hyperparameters for the Decision Tree, such as 'max_depth' which defines the maximum depth of the tree, 'min_samples_split' which defines the minimum number of samples required to split an internal node, and 'min_samples_leaf', which define the minimum number of samples required to be at a leaf node. All these hyperparameters collectively determine the structure and complexity of the tree. The best hyperparameters are chosen based on cross-validation.

4.5 Model Evaluation

For the evaluation of the selected machine learning models, the metrics selected are based on previously conducted studies in this domain, namely, Accuracy, Sensitivity or recall, Specificity, Precision, F1 score and AUC. Additionally, we employed another metric called the Cohen's Kappa. We used this to measure the agreement between the actual and predicted labels while accounting for chance agreement, as it also takes into account the possibility of the agreement occurring by chance. Cohen's Kappa is particularly useful when evaluating the performance of classifiers in situations where there might be an imbalance in the distribution of classes. Since all three selected datasets had the issue of imbalance, we opted to use this for evaluating the performance of our classifiers to ensure a more accurate assessment of their effectiveness. A visual representation of the complete methodology is provided in Figure 4

4.6 Conclusion

This chapter details the methodology for developing and evaluating models to predict suicide ideation risk among medical students. The process started with collecting publicly available datasets focused on LMIC, followed by data cleaning and preprocessing, which included handling missing values, converting categorical data to numeric using label encoding, and applying min-max normalization. These steps were crucial for preparing the datasets for effective model training. SMOTE was used as the data balancing technique to address class imbalance. This

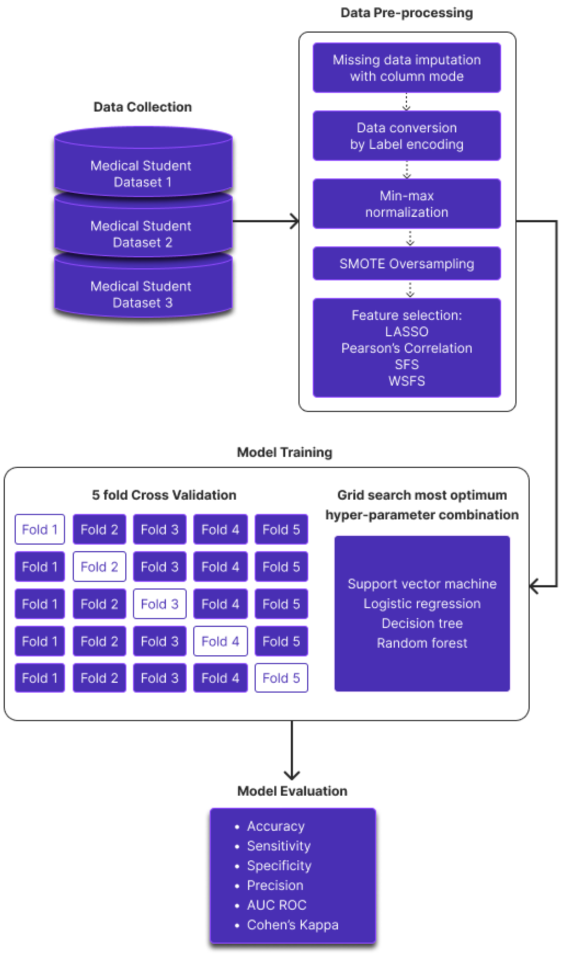

Figure 4: Complete methodology of the prediction model.

ensured that the machine learning models could perform well on both the minority and majority classes, thereby improving the generalization of the models. Feature selection was done using four different methods, LASSO, Sequential Feature Selection, Weighted Sum Feature Selection and Pearson's correlation coefficient to identify the most relevant features for predicting suicide ideation. The methodology focused on training and evaluating four machine learning algorithms: Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF). The hyperparameters of these models were optimized using grid search combined with 5-fold cross-validation, ensuring that the models were fine-tuned to achieve the best performance. The evaluation metrics used for this study are: accuracy, sensitivity, specificity, precision, F1 score, area under the ROC curve (AUC), and Cohen's Kappa.. The results of the predictive models will be discussed in detail in the next chapter.

Chapter 5

Result Analysis

This chapter evaluates the performance of a suicide ideation prediction model developed by applying several machine learning algorithms on 3 medical student datasets. Additionally, the influence of various feature selection methods on these models are explored, with the aim of determining the most effective combinations of these approaches to enhance the accuracy and reliability of a suicide ideation prediction model.

Initially, Lasso was chosen as the feature selection method due to its success in similar studies. It was applied on all three datasets to select the relevant features before training the machine learning algorithms. Results show that Random Forest achieved the best performance, with a mean accuracy of 91%. Decision Trees performed well, with a mean accuracy of 86%. Logistic Regression and SVM also showed acceptable metrics, with mean accuracy of 80%. The reliability of these results is further supported by other metrics, particularly Cohen's Kappa, where Random Forest consistently achieved higher values, indicating stronger agreement and model performance. The mean model performance of all datasets using LASSO is shown in figure 5 represented as a bar graph.

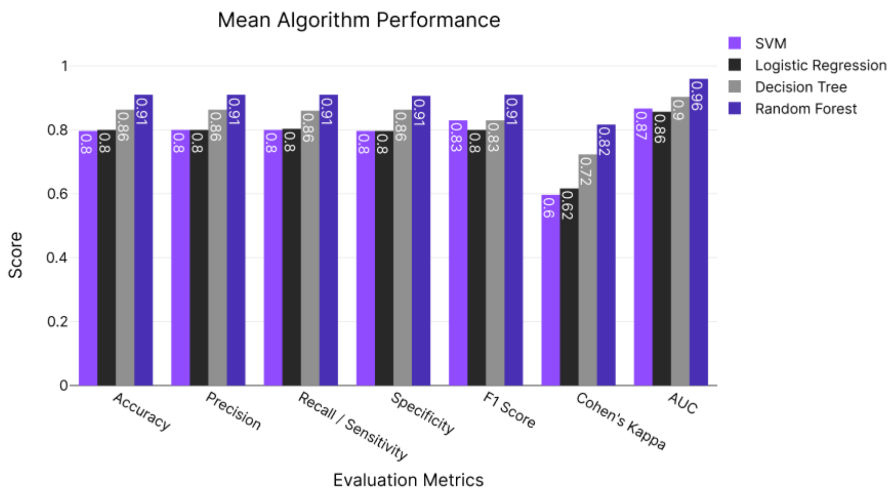

Figure 5: Mean model performance using Lasso feature selection.

Additionally, other feature selection techniques were explored to observe the effect on the model performance, namely Pearson's Correlation Coefficient (PCC), Sequential Forward Selection (SFS) and the Weighted Sum Feature Selection (WSFS). Among these techniques, WSFS significantly improved the performance of the algorithms and datasets. It delivered with the highest accuracy, precision, and AUC, particularly with Random Forest, where it achieved 97% accuracy and 99% AUC on Dataset A, and 92% accuracy along with 98% AUC on Dataset C. One exception occurred when the model had been developed using Dataset B, where LASSO outperformed WSFS. Conversely, WSFS benefitted other algorithms like Decision Trees and Logistic Regression that showed strong and stable performance due to higher Kappa scores. Pearson's Correlation Coefficient (PCC) also resulted in moderate performance. While Random Forest manages to produce acceptable results with mean accuracy of 85%, the results are notably weaker than with WSFS or Lasso. SFS achieved acceptable results with Random Forest (mean accuracy 85%) and Decision Tree (mean accuracy 79%), but showed poor performance for SVM and Logistic Regression, with a mean accuracy of 69% each. Specificity values fluctuated for both the algorithms, but the f1 scores were similar, though Logistic Regression often edged out SVM, achieving better balance between precision and recall. Furthermore, both SFS and PCC produced low Kappa scores, indicating poorer agreement between the predicted and actual classifications. Figure 6 and 7 help to visualize and compare the mean performance of each feature selection.

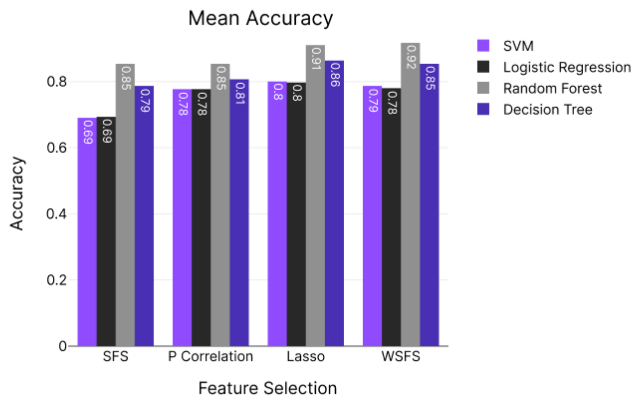

Figure 6: Bar graph comparison of all the feature selection methods and the algorithms.

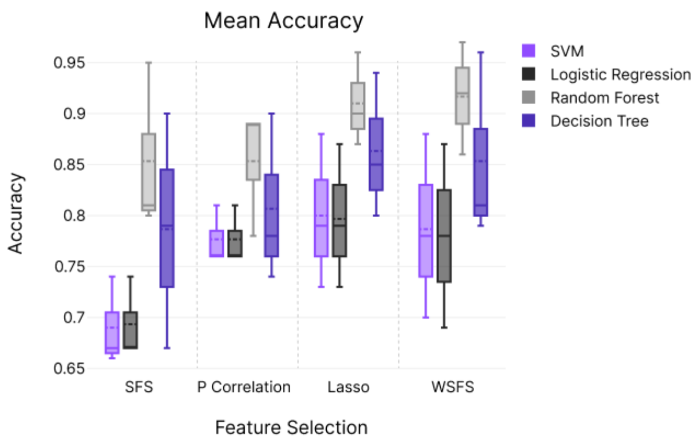

Figure 7: Box plot diagram for comparison of all the feature selection methods and the algorithms.

In general, Random Forest gave the best results followed by Decision Trees while Logistic Regression and SVM showed competing performance giving almost same accuracy readings for most of the cases. For all three datasets, Random Forest showed the best performing model with features selected using WSFS in Dataset A and C (see Table 10 and 11), and using LASSO in Dataset B (see Table 12). Thus, comparing the overall performance, Random Forest with features selected using WSFS performed the best with a mean accuracy of 92% backed by Cohen's Kappa value of 0.84. Table 13 summarizes the results of all developed prediction models and their evaluations, sorted in descending order as the best performing model.

Table 10: Best features selected by WSFS in Dataset A.

Gender	Family suicide history
Year of study	Chronic psychological sufferings since last year
Family monthly income	Post-traumatic events since last year
Psychoactive substance user	Suicidal Attempt Lifetime

Table 11: Best features selected by WSFS in Dataset C.

Hometown	Allergy	Stress due to Relationship with one's beloved
Enough Income	Peer pressure alcohol	Stress due to Relationship with family
Live with (House members)	Stress alcohol	Stress due to Relationship with instructor
Living place	Treating psychiatric symptoms	Stress due to Physical health
Medical comorbidity	Difficult social relation	Stress due to Mental health
Apply to wrong faculty	Enough grade	Stress due to Living place
Emotional exhaustion	Bore to learn	Stress due to Health: Physical and mental
Depersonalization	Playing game	Stress due to financial problem
Personal efficacy	Stress due to Living place, future, political issue, family's health	Stress due to Relationship with friends

Table 12: Best features selected by LASSO in Dataset B.

Family Problems	Punishment Feelings	Changes in Sleeping Pattern
Kind of drug addicted to	Self Dislike	Tiredness or Fatigue
Past Failure	Worthlessness	Loss of Interest in Sex

Table 13: Comprehensive Performance Metrics Comparison Across Datasets and Algorithms

Algo.	Dataset	Feature Selection	Acc.	Prec.	Recall/ Sens.	Spec.	F1	Cohen's Kappa	AUC
RF	1	WSFS	0.97	0.97	0.98	0.97	0.97	0.95	0.99
RF	1	Lasso	0.96	0.97	0.95	0.97	0.96	0.92	0.99
DT	1	WSFS	0.96	0.97	0.96	0.97	0.96	0.92	0.97
RF	1	SFS	0.95	0.96	0.95	0.96	0.95	0.91	0.97
DT	1	Lasso	0.94	0.96	0.93	0.96	0.84	0.89	0.97
DT	1	P Corr.	0.90	0.93	0.86	0.93	0.89	0.80	0.96
DT	1	SFS	0.90	0.90	0.90	0.90	0.90	0.80	0.94
RF	1	P Corr.	0.89	0.92	0.86	0.92	0.89	0.79	0.95
LR	1	Lasso	0.88	0.88	0.89	0.88	0.88	0.77	0.92
LR	1	WSFS	0.88	0.87	0.89	0.86	0.88	0.75	0.93
SVM	1	Lasso	0.87	0.88	0.87	0.88	0.97	0.75	0.94
SVM	1	WSFS	0.87	0.87	0.88	0.86	0.87	0.75	0.94
LR	1	P Corr.	0.81	0.88	0.73	0.90	0.80	0.63	0.86
SVM	1	P Corr.	0.81	0.87	0.73	0.89	0.80	0.63	0.89
SVM	1	SFS	0.67	0.67	0.68	0.66	0.67	0.34	0.76
LR	1	SFS	0.66	0.66	0.68	0.65	0.67	0.33	0.75
RF	2	Lasso	0.87	0.87	0.87	0.87	0.87	0.74	0.93
RF	2	WSFS	0.86	0.88	0.84	0.88	0.86	0.73	0.93
DT	2	Lasso	0.85	0.84	0.85	0.84	0.85	0.69	0.88
RF	2	SFS	0.81	0.86	0.74	0.88	0.80	0.63	0.86
DT	2	SFS	0.79	0.87	0.69	0.90	0.77	0.59	0.83
DT	2	WSFS	0.79	0.80	0.78	0.80	0.79	0.58	0.84
LR	2	Lasso	0.79	0.80	0.76	0.81	0.78	0.58	0.86
SVM	2	Lasso	0.79	0.80	0.77	0.81	0.78	0.58	0.86
RF	2	P Corr.	0.78	0.64	0.42	0.91	0.50	0.37	0.80
LR	2	P Corr.	0.76	0.56	0.48	0.86	0.52	0.35	0.80
SVM	2	P Corr.	0.76	0.59	0.42	0.89	0.49	0.34	0.78
DT	2	P Corr.	0.74	0.54	0.40	0.87	0.46	0.30	0.71
LR	2	SFS	0.74	0.77	0.68	0.80	0.72	0.48	0.79
SVM	2	SFS	0.74	0.78	0.67	0.81	0.72	0.48	0.80
LR	2	WSFS	0.70	0.70	0.70	0.70	0.70	0.40	0.80
SVM	2	WSFS	0.69	0.69	0.70	0.68	0.70	0.38	0.79
RF	3	WSFS	0.92	0.90	0.94	0.90	0.92	0.84	0.98
RF	3	Lasso	0.90	0.89	0.91	0.88	0.90	0.79	0.96
RF	3	P Corr.	0.89	0.87	0.91	0.86	0.89	0.77	0.96
DT	3	WSFS	0.81	0.79	0.83	0.78	0.80	0.61	0.85
RF	3	SFS	0.80	0.76	0.87	0.72	0.81	0.59	0.90
DT	3	Lasso	0.80	0.79	0.80	0.79	0.80	0.59	0.86

Continued on next page

Table 13 – Continued from previous page

Algo.	**Dataset**	**Feature Selection**	**Acc.**	**Prec.**	**Recall/ Sens.**	**Spec.**	**F1**	**Cohen's Kappa**	**AUC**
DT	3	P Corr.	0.78	0.76	0.82	0.73	0.79	0.56	0.83
LR	3	WSFS	0.78	0.74	0.86	0.70	0.80	0.56	0.83
SVM	3	WSFS	0.78	0.74	0.86	0.69	0.79	0.55	0.81
LR	3	P Corr.	0.76	0.76	0.75	0.76	0.76	0.52	0.83
SVM	3	P Corr.	0.76	0.76	0.77	0.75	0.76	0.52	0.82
LR	3	Lasso	0.73	0.72	0.76	0.70	0.74	0.50	0.79
SVM	3	Lasso	0.73	0.72	0.76	0.70	0.74	0.46	0.80
DT	3	SFS	0.67	0.64	0.79	0.56	0.71	0.35	0.75
LR	3	SFS	0.67	0.66	0.69	0.64	0.68	0.33	0.72
SVM	3	SFS	0.67	0.66	0.69	0.65	0.67	0.33	0.72

Moreover, a secondary experiment was conducted to validate the model, where two datasets were combined and validated by the third. The three datasets were analyzed to identify common features, with the resulting features presented below:

- Suicidal Thoughts
- Family History of depression
- Year of Study
- Monthly Family Income
- Addicted to Drugs
- Type of Family
- Gender

After identifying the common features, the Leave One Out (LOO) approach was taken where two datasets were combined to be one training set while the remaining one dataset was used for testing. This process was repeated three times, with each dataset taking a turn as the test set.
First, a model with medical student datasets from Bangladesh (Dataset B) and from Thailand (Dataset C) were combined and trained to try to predict suicide risks among Dental students from Bangladesh (Dataset A). The model showed poor performance overall because Cohen's Kappa gave a negative result for both logistic regression and Random Forest. Table 14 shows the evaluation of this model.

Table 14: Trained with Medical Student Datasets from Bangladesh (Dataset B) and Thailand (Dataset C). Tested with Dental Student Dataset from Bangladesh (Dataset A)

Dataset A	Acc.	Prec.	Recall / Sensitivity	Spec.	F1 Score	Cohen's Kappa	AUC
LR	0.29	0.02	0.53	0.29	0.05	-0.02	0.51
RF	0.6	0.27	0.33	0.61	0.05	-0.01	0.4
SVM	0.44	0.38	0.67	0.44	0.07	0.01	0.57
DT	0.53	0.04	0.6	0.53	0.08	0.02	0.51

Next, the model with dental student dataset from Bangladesh (Dataset A) and medical student dataset from Thailand (Dataset C) were combined and trained to try to predict suicide risks among medical students from Bangladesh (Dataset B). This model also showed below par performance overall because the value of Cohen's Kappa for all the algorithms were close to 0. Table 15 showcases the evaluation of this model.

Table 15: Trained with Dental student dataset from Bangladesh (Dataset A) and Medical Student Dataset from Thailand (Dataset C). Tested with Medical Student Dataset from Bangladesh (Dataset B)

Dataset B	Acc.	Prec.	Recall / Sensitivity	Spec.	F1 Score	Cohen's Kappa	AUC
LR	0.47	0.3	0.69	0.39	0.42	0.06	0.62
RF	0.43	0.28	0.68	0.34	0.39	0.01	0.45
SVM	0.5	0.25	0.42	0.53	0.31	-0.05	0.44
DT	0.45	0.28	0.63	0.38	0.39	0.01	0.49

Finally, the model with dental student dataset from Bangladesh (Dataset A) and medical student dataset from Bangladesh (Dataset B) were combined and trained to try to predict suicide risks among medical students from Thailand (Dataset C). This model also failed to show good performance given Random Forest and Decision Trees showcased negative Kappa values. Although SVM had an accuracy of 77%, the Kappa value was 0.03. Table 16 showcases the evaluation of this model.

Table 16: Trained with Dental student dataset from Bangladesh (Dataset A) + Medical Student Dataset from Bangladesh (Dataset B). Tested with Medical Student Dataset from Thailand (Dataset C)

Dataset C	Acc.	Prec.	Recall / Sensitivity	Spec.	F1 Score	Cohen's Kappa	AUC
LR	0.69	0.14	0.3	0.75	0.19	0.03	0.51
RF	0.49	0.12	0.5	0.49	0.19	-0.003	0.45
SVM	0.77	0.15	0.17	0.86	0.16	0.03	0.49
DT	0.51	0.1	0.38	0.53	0.16	-0.04	0.45

Chapter 6

Conclusion

Since the research on this topic is scarce, the contents of this chapter has been continued separately to provide a clearer picture of suicidal ideation among medical students. In this study, three datasets were accumulated which resulted in some specific outcomes which will be analyzed in this chapter. The details of the research findings including the performance of the prediction models, which of them was more effective in predicting suicide risks, variability, and analyzing the performance metrics will be discussed. Additionally, the rationale for selecting the particular techniques of feature selection featured in this study will be analyzed here. Moreover, the results of the combination of the three datasets for predicting suicidal outcomes will be discussed here. Several limitations and challenges have been encountered that have either hindered or slowed down the processes for this research. Thus, more time was required to accomplish this research, which altered an initially planned approach. Such limitations and challenges had to be overcome through specific methods and approaches, which will be discussed in this chapter. Lastly, on the basis of the research scope of this study highlighted in the research objectives above, future scope and recommendation will be given to the research community to conduct further research to extend this topic.

6.1 Model performance

Among the implemented algorithms, Random Forest generally exhibited better performance in most cases. It was highly effective in handling the complexities of the selected features in all datasets with high accuracy and Cohen's Kappa scores, making it suitable for medical setups. Furthermore, both recall and sensitivity are balanced, which indicates that the model is effective in identifying true positives (sensitivity) while maintaining a strong ability to avoid false negatives (high recall). The AUC scores, which consistently approach 0.99 across different feature selection methods, demonstrate the Random Forest's excellent ability to differentiate between classes. Its ensemble approach along with tree based structure employs bagging that reduces overfitting and allows it to perform well even with varied and imbalanced data, leading to higher accuracy and reliability.

Decision Tree also achieved strong performance with high accuracy and Kappa scores in maximum cases. It achieved high precision, recall, and F1 scores in the simpler datasets, with accuracy reaching up to 0.96 and Cohen's Kappa as high as 0.92, indicating effective classification and agreement between predicted and actual labels. However, Decision Trees are prone to overfitting, which impacts their performance on more complex datasets. For instance, the lowest accuracy was found to be as low as 0.67 and the Cohen's Kappa to be as low as 0.35. Additionally, even though the AUC scores remain relatively high, the model's performance declined with complex datasets and less effective feature selection methods. This explains that Decision Trees are powerful in simpler settings such as in Dataset A with in predicting suicide ideation when features were selected using PCC. But, their generalizability issues make them less robust compared to ensemble methods like Random Forest.

In the case of Logistic Regression and Support Vector Machine provided, they barely provided acceptable performance overall, with accuracy generally around 0.87 to 0.88. SVM however showed more variability in precision and recall. It captured higher true positives along with false positives. Logistic regression maintained a more consistent balance comparatively. Cohen's Kappa scores also suggested moderate agreement, particularly in cases of lower accuracy. These algorithms rely on simpler, linear relationships, and given the nature of the selected features, they were not able to capture the complexity of suicidal ideation.

6.2 Effects of Feature Selection method used

The four feature selection methods demonstrated a mixed result based on the algorithms and datasets. LASSO consistently delivered strong results across all algorithms, especially for Random Forest and Decision Tree, where it maintained accuracy above 85% across most datasets. Its ability to minimize overfitting by penalizing less relevant features likely contributed to its high precision and specificity, as demonstrated in the SVM with Dataset A (0.88 precision, 0.88 specificity) and Random Forest with Dataset A (0.97 accuracy, 0.97 specificity). Lasso's regularization also likely enhanced model generalization, evident from the stable performance metrics across different datasets.

As for WSFS, it performed comparably well, achieving the highest overall performance through Dataset A. The weighting features based on importance contributed to the robustness of WSFS, making it very effective for ensemble models like Random Forest. However, for SVM, WSFS displayed varying results, ranging from high performance in Dataset A (0.87 accuracy, 0.87 precision) to a significant drop in Dataset B (0.69 accuracy, 0.69 precision). This shows that WSFS may be more sensitive to dataset structure or feature redundancy. SFS on the other hand was beneficial for simpler models, but showed limited effectiveness with more complex datasets and algorithms. Its performance with SVM was generally lower compared to Lasso or WSFS, as seen in Dataset C (0.67 accuracy, 0.66 precision). SFS often resulted in lower recall and F1 scores, suggesting that its greedy selection process, which chooses features sequentially based on individual contribution, might overlook combinations of features that work synergistically. This can explain why SFS performed better with Random Forest on Dataset B (0.81 accuracy) than SVM or Logistic Regression.

Pearson's Correlation Coefficient struggled in several cases, especially with algorithms like Decision Tree and Logistic Regression on Dataset B (0.54 accuracy, 0.40 recall for Decision Tree). The limitation of Pearson's correlation in capturing nonlinear relationships between features likely hindered its effectiveness with these models. Despite this, Pearson's Correlation did show reasonable performance with Random Forest and SVM in some datasets (e.g., Random Forest on Dataset C with 0.89 accuracy), highlighting its utility when linear correlations dominate the feature space.

All in all, it can be concluded that Lasso proved to be the most reliable feature selection method across different algorithms and datasets when it comes to handling both complex and simple data structures. WSFS proved to be highly effective with ensemble methods and on more structured datasets, while SFS and Pearson's Correlation are better suited for smaller or simpler feature spaces.

6.3 Data type and predictability

The three datasets in this study cover a variety of demographic, educational, family, lifestyle, and mental health factors.
In Dataset A, regardless of the feature selection technique used, standalone and ensemble tree based algorithms (Decision Trees and Random Forest respectively) effectively managed the complex, non-linear relationships by properly identifying and utilizing key features such as chronic physical disease, psychoactive substance use etc. The same cannot be said for SVM and Logistic Regression since they performed well only after being paired with LASSO and WSFS. Here, both LASSO and WSFS focused on selecting the most informative variables, such as 'Family mental illness history' and 'Chronic psychological sufferings since last year,' while eliminating less relevant features like 'Religion' and 'Cigarette-smoker.' As a result, dimensionality and noise were reduced, which enhanced model accuracy and generalization. As for Pearson's Correlation and SFS, they could not effectively prioritize the most impactful features due to their linearity, leading to suboptimal model performance.
In Dataset B, which is rich in symptom-related variables (e.g., sadness, loss of pleasure) and includes demographic and educational data, LASSO consistently showed strong performance with various algorithms. Lasso's ability to reduce dimensionality and focus on the most relevant variables proved beneficial, particularly in models using Logistic Regression and Decision Trees. With Logistic Regression, Lasso helped isolate key predictors of Suicidal Thoughts, achieving high accuracy and precision. WSFS also provided notable performance improvements, especially with Random Forests, which effectively handled the wide range of symptom variables by weighting them appropriately. SFS, on the other hand, struggled in this dataset due to its sequential approach, which may have missed critical feature interactions related to symptoms and mental health. Pearson's Correlation was limited in capturing the non-linear relationships within the symptom variables, which led to less effective performance in models like Logistic Regression.
For Dataset C, which encompasses a broad array of stress-related, lifestyle, and medical history variables, Lasso was again highly effective, particularly in combination with SVM and Decision Trees. The regularization technique of Lasso helped manage the extensive feature set by selecting the most impactful features, contributing to high accuracy and precision with SVM and Decision Trees. WSFS proved beneficial with Random Forests, as it effectively handled the dataset's complexity by weighting diverse features such as financial stress and emotional exhaustion. SFS faced challenges in this dataset, as its sequential feature selection was less adept at managing the interactions among a wide variety of features, leading to lower performance metrics. Pearson's Correlation struggled with capturing the non-linear and complex relationships among stressors and lifestyle factors, which impacted its effectiveness in models like Logistic Regression and Decision Trees.
The interaction between feature selection methods and machine learning algorithms significantly impacted model performance across datasets. Lasso consistently demonstrated its strength in handling high-dimensional data by focusing on relevant features, while WSFS excelled in scenarios requiring weighted feature importance. SFS and Pearson's Correlation faced limitations in managing complex, non-linear relationships and high-dimensional feature sets. The choice of algorithm also played a crucial role, with Decision Trees and Random Forests often outperforming SVM and Logistic Regression in handling intricate feature interactions and non-linear relationships within

the datasets.
Furthermore, while evaluating the performance of feature selection techniques, it was observed that WSFS (Weighted Sum Feature Selection) achieves the best accuracy for Dataset A and Dataset C. Such results suggest that the technique was particularly suitable for identifying potential feature sets that effectively captured the details of each dataset, thereby enhancing the model's predictive power. It can be concluded that since Dataset A had fewer features compared to Dataset C, it gave the best results, which may indicate a case of overfitting. However, Dataset C had a more diverse number of data categories, along with a vast number of symptom based data, mental health history data and lifestyle data extracted for each sample. WSFS applied on a dataset with this type of complex attribute selects features more accurately. On the other hand, for Dataset B, the best algorithm originated from utilizing LASSO for feature selection in each experiment. The reason for the success in this particular case may have been because of the function that is able to handle variable selection and regularization. It can also be due to the fact that LASSO had more symptom based data compared to the other categories of data to deal with.
For the poorer performing cases, a varied pattern of performance was found across different datasets. For Dataset A, it was observed that Logistic Regression and SVM algorithms exhibited the worst performance when using features selected by the forward Sequential Forward Selection (SFS). This output may be due to the fact that the particular subset of features chosen by SFS did not align well with the predictive needs of these models for this dataset, leading to suboptimal results. Additionally, the sequential nature of SFS can lead to the inclusion of features that, while individually significant, collectively introduce redundancy or multicollinearity, further degrading the performance of these linear classifiers. This mismatch between the selected features and the capabilities of Logistic Regression and SVM likely results in their poor performance on Dataset A. Conversely, for Dataset B, Logistic Regression and SVM classifiers showed the lowest accuracy levels when WSFS was employed to select features. WSFS selects features that favor non-linear interactions and complex models, which these linear classifiers cannot effectively leverage. This inconsistency highlights how different feature selection methods can significantly impact the efficacy of algorithms depending on the dataset in question. Also, Decision Tree and Random Forest models showed the poorest performance in dataset B when features were selected using Pearson's Correlation Coefficient. This outcome suggests that Pearson's Correlation Coefficient might not be an effective feature selection method for these particular tree-based algorithms, potentially due to its focus on linear relationships, which might not capture the complex interactions required for these models. This approach may overlook important nonlinear interactions, which are crucial for these models. Additionally, Dataset B contains a large number of symptom features (20), which could be highly correlated with each other, leading to redundancy. Pearson's correlation might also inadequately handle categorical features from demographic, lifestyle, economical, and educational categories, resulting in a suboptimal feature set that fails to enhance model performance.
When the features were selected using SFS on Dataset C, all algorithms performed poorly. This widespread inefficacy indicates a broader issue with the suitability of SFS for feature selection within this dataset, affecting a diverse range of models. The reason for such results may be due to the high number of features, particularly in the lifestyle (12), medical history (7), and mental health history (15) categories. This abundance increases the likelihood of redundancy and multicollinearity, complicating the feature selection process. SFS, which adds features one by one based on their individual contribution to model performance, may struggle to effectively discern the most informative features amidst the large, diverse, and potentially correlated feature set, leading to suboptimal model performance.

6.4 Combination of datasets

In the second approach, the common attributes of the datasets were combined to check if medical student data from Bangladesh and overseas can be merged to create a good prediction model. The result of all three models showed poor performance. Specifically, negative Cohen's Kappa values of the logistic regression and Random Forest models imply that the models poorly generalize the suicide risks among Dental students from Bangladesh based on combined datasets of Medical students from Bangladesh and Thailand; the poor alignment of predicted and actual labels can also be essential for poor performance. Some reasons for this poor quality could be the lack of some features from the combined datasets (Dataset B and C) to the target dataset (Dataset A) and potential data imbalance or class distribution issues in Dataset A.
The model combining the dental student dataset from Bangladesh (Dataset A) and the medical student dataset from Thailand (Dataset C) to predict suicide risks among medical students from Bangladesh (Dataset B) showed poor performance indicated by Cohen's Kappa values close to 0 across all algorithms. Potential explanations may cover such areas as the small overlap of features of the general dental student population and medical student population respectively to provide adequate learning of predictors for the suicide risks among the latter. In the same way, dental and medical students might have been unsuitable samples for crossover because contextual, demographic, or risk factors could have hampered the models' applicability further. These create a necessity to ensure that datasets used in model development are compatible, especially when developing models across populations and domains of endeavor.
The combined model using dental student dataset from Bangladesh (Dataset A) and medical student dataset from Bangladesh (Dataset B) to predict suicide risks among medical students from Thailand (Dataset C) resulted in yet another poor performance, identified by negative Cohen's Kappa values that was particular to both Random Forest and Decision Trees. The negative Kappa values mean that there was a higher disagreement between the predicted and the actual labels, thus providing evidence that the models' decisions were inaccurate and arbitrary. The issues may be arising from major differences between the training set (Dental & Medical students from Bangladesh) and the target set (Medical students from Thailand).

6.5 Comparison with existing models

Research regarding suicide prediction models of medical students have been circulating for a while now. However, no articles regarding suicide have been formally introduced in the research field of Bangladesh, let alone medical students. Therefore, the results of this study will be compared to similar studies related to suicide prediction from a global perspective.
Among all the studies in the literature regarding suicide risk prediction, our model performed the best, comparatively. When WSFS was run on Dataset A to select features, Random Forest provided an accuracy of 0.97 and an AUC of 0.99. Zuromski et al. predicted suicide attempts among soldiers, achieving an AUC of 0.77 [113] Rus Prelog et al. achieved an AUC of 0.83 while [112]. Su et al. ran the prediction model on two datasets, where an accuracy of 0.81 was achieved on the self harm dataset and 0.88 was achieved on the suicide attempt dataset [120]. Wang et al. achieved an AUC of 0.901 while trying to predict suicidal ideation among middle-aged populations [119], thus showing that our model had the best accuracy comparatively.

6.6 Limitation

Although collection of the three datasets were completed successfully, for the purpose of this study, there were fewer items in all the dataset resulting in a class imbalance. This affected the credibility and validity of our work, as the data distributions for the classes were not balanced, which limited our ability to generalize our conclusion. Additionally, it was evident that the feature variables in these datasets were not standardized and presented other complications. Issues related to data formatting and scaling, thus, rendered the information as not very compatible with each other for effective analysis and had a negative impact on the results of the combination model. These limitations highlight the need for data quality and preprocessing in machine learning research, especially when the resources that are available for data collection and preparation are limited.
The dataset A is prone to bias since no validation of any sort was conducted, nor were any physicians or psychiatrists appointed to validate the items of the dataset to be collected. Although datasets B and C are research backed that use scale data as well, the datasets still remain limited. Furthermore, data from LMIC is a bit scarce when compared to data being used in already implemented real world applications, assessments and diagnostic tools. Even those included within the public domain from lower middle-income countries such as Bangladesh do not necessarily encapsulate the relevant socio-cultural factors required. This lack of data that is unique to the respondent's culture posed to be a limitation in this study, which in turn also limited the extensiveness of the impact on the phenomena of interest. We initially thought of curating a dataset that would have been appropriate for our research, but due to resource and time constraints, we proceeded with available public datasets. Nonetheless, this research offers new insights into the possibilities and limitations of employing pre-existing datasets from LMICs.
Advanced techniques like deep learning and large language models (LLMs) offer significant potential. However, their complexity and high computational demand was a limitation in this study, which did not allow moving forward in this path due to resource constraints. Therefore, well-established and validated techniques were deliberately used for the problem at hand, ensuring robust and actionable outcomes at the cost of limiting our ability to capture complex relationships in the data.
This study is first and foremost about an attempt to design a machine learning model for probable suicide risk assessment among Bangladeshi students with the help of multiple surveys datasets. The focus is based on the possibility of using AI technologies in managing mental health challenges specific to this population. For the purpose of increasing the efficiency of the predictive models, the study pays special attention to the aspect of classification of data depending on their cultural relevance. Therefore, unlike similar studies that offer general solutions, the research presented herein targets Bangladeshi medical students with the idea of offering specific solutions.
To overcome the issue of bias due to oversampling and to develop a better quality of model, balancing the dataset while carrying out extensive data preprocessing techniques was attempted. Additionally, in order to create a medical suicide risk prediction for the identified students, four machine learning algorithms were used: Decision Trees, Logistic Regression, Random Forest, and SVM. This kind of approach helped to reduce the subjectivity of the analysis and also contributed to the increase in the stability of the predictive indicators. Having reviewed these algorithms, the best model for estimating suicide risks in this regard is determined. The outcome revealed the promise of ML in providing insights and solutions to mental health concerns of Bangladeshi learners.
Furthermore, evaluation of the performance of four feature selection algorithms across various

datasets has been conducted to assess their impact on predictive accuracy. These algorithms included Sequential Forward Selection Forward (SFS), Weighted Sum Feature Selection (WSFS), Lasso regression, and Pearson's correlation coefficient. Our study focused on finding out which of the methods can be used to improve the feature set and improve the predictive suicide risk model for Bangladeshi medical students. In this way, optimization of the feature selection process of the model to identify the most robust predictors of suicide risk in the Bangladeshi student population can be found. This makes the model more accurate and adaptable to solve problems in real life situations.

Further, to improve the prediction accuracy of suicide risk, the possibility of merging the three datasets was examined. A detailed exploration on how the combination of these datasets affected the performance of the given predictive models is also provided. Lastly, the detected studies were critically reviewed towards determining whether the proposed model offers a proper way of identifying and evaluating all the significant risk factors of suicide amongst the Bangladeshi medical students. This investigation was intended to determine if a multisite approach could provide a better identification of suicide risk factors unique to Bangladeshi students. In this way, relevant findings are provided that would help in designing specific approaches and actions that would support positive mental health amongst young people.

Future research should prioritize broadening the scope and enhancing the effectiveness of predictive models for assessing suicide risks in Bangladesh. While locally developed risk tools and scales can be used for risk assessment, their accuracy has not been well validated. Through the use of this research, an application to assess the mental state of medical students can be developed, which will be used for identifying individuals at risk of suicide. Additionally, those selected students can be allocated to appropriate psychiatrists based on their specific needs. Also, an early intervention tool for suicide prevention can be created to treat at-risk medical students to reduce suicide rates in Bangladesh.

References

[1] Abraham Haileamlak. Mental health disorders: the deserted illnesses. *Ethiopian journal of health sciences*, 27(1):1–2, 2017.

[2] Mark Rice-Oxley. Mental illness: is there really a global epidemic. *The Guardian*, 3:6–7, 2019.

[3] Kim Storrie, Kathy Ahern, and Anthony Tuckett. A systematic review: students with mental health problems—a growing problem. *International journal of nursing practice*, 16(1):1–6, 2010.

[4] Anja Susanne Dessauvagie, Hoang-Minh Dang, Thi Anh Thu Nguyen, and Gunter Groen. Mental health of university students in southeastern asia: a systematic review. *Asia Pacific Journal of Public Health*, 34(2-3):172–181, 2022.

[5] Tally Moses. Being treated differently: Stigma experiences with family, peers, and school staff among adolescents with mental health disorders. *Social science & medicine*, 70(7):985–993, 2010.

[6] Bethany A Rallis, Laura Hungerford, and Chris Flynn. A wave of depression: Implications for college student mental health. *Journal of College Student Mental Health*, 38(3):541–551, 2024.

[7] Randy P Auerbach, Philippe Mortier, Ronny Bruffaerts, Jordi Alonso, Corina Benjet, Pim Cuijpers, Koen Demyttenaere, David D Ebert, Jennifer Greif Green, Penelope Hasking, et al. Who world mental health surveys international college student project: Prevalence and distribution of mental disorders. *Journal of abnormal psychology*, 127(7):623, 2018.

[8] Bahr Weiss, Victoria Khanh Ngo, Hoang-Minh Dang, Amie Pollack, Lam T Trung, Cong V Tran, Nam T Tran, David L Sang, and Khanh N Do. A model for sustainable development of child mental health infrastructure in the lmic world: Vietnam as a case example. *International Perspectives in Psychology*, 1(1):63–77, 2012.

[9] Katie S Dawson, Richard A Bryant, Melissa Harper, Alvin Kuowei Tay, Atif Rahman, Alison Schafer, and Mark Van Ommeren. Problem management plus (pm+): a who transdiagnostic psychological intervention for common mental health problems. *World Psychiatry*, 14(3):354, 2015.

[10] Md Omar Faruk. Community-based mental health services in bangladesh: Prospects and challenges. *World Social Psychiatry*, 4(3):187–192, 2022.

[11] Tunvir Ahamed Shohel, Nishad Nasrin, Fariha Farjana, Taufiq-E-Ahmed Shovo, Aisha Rahman Asha, Morsheda Akter Heme, Ashraful Islam, Pranto Paul, and Md Tanvir Hossain. 'he was a brilliant student but became mad like his grandfather': an exploratory investigation on the social perception and stigma against individuals living with mental health problems in bangladesh. *BMC psychiatry*, 22(1):702, 2022.

[12] Anwar Islam and Tuhin Biswas. Mental health and the health system in bangladesh: situation analysis of a neglected domain. *Am J Psychiatry Neurosci*, 3(4):57–62, 2015.

[13] Anaise Williams, Malabika Sarker, and Syeda Tahmina Ferdous. Cultural attitudes toward postpartum depression in dhaka, bangladesh. *Medical anthropology*, 37(3):194–205, 2018.

[14] Md Abu Bakar Siddique, Musaddiqur Rahman Ovi, Tanvir Ahammed, Muhammad Abdul Baker Chowdhury, and Md Jamal Uddin. Mental health knowledge and awareness among university students in bangladesh. *Heliyon*, 8(10), 2022.

[15] SM Yasir Arafat. Suicide in bangladesh: a mini review. *Suicide*, 3, 2014.

[16] Shagoofa Rakhshanda, Cinderella Akbar Mayaboti, Minhazul Abedin, Farah Naz Rahman, AKM Fazlur Rahman, Salim Mahmud Chowdhury, and Saidur Rahman Mashreky. 443 changes in suicide in bangladesh: Findings from bhis 2003 and 2016. *Injury prevention*, 28(Suppl 2):A67–A67, 2022.

[17] Tanjir Rashid Soron and Sheikh Mohammed Shariful Islam. Suicide on facebook-the tales of unnoticed departure in bangladesh. *Global Mental Health*, 7:e12, 2020.

[18] M Rasheduzzaman, Firoj Al-Mamun, Ismail Hosen, Tahmina Akter, Moazzem Hossain, Mark D Griffiths, and Mohammed A Mamun. Suicidal behaviors among bangladeshi university students: Prevalence and risk factors. *PLoS one*, 17(1):e0262006, 2022.

[19] Morton M Silverman and Alan L Berman. Suicide risk assessment and risk formulation part i: A focus on suicide ideation in assessing suicide risk. *Suicide and Life-Threatening Behavior*, 44(4):420–431, 2014.

[20] Morton M Silverman. Suicide risk assessment and suicide risk formulation: essential components of the therapeutic risk management model. *Journal of Psychiatric Practice®*, 20(5):373–378, 2014.

[21] Emilie H Picard and Barry Rosenfeld. How clinicians incorporate suicide risk factors into suicide risk assessment. *Crisis*, 2020.

[22] Meizhen Lv, Ang Li, Tianli Liu, and Tingshao Zhu. Creating a chinese suicide dictionary for identifying suicide risk on social media. *PeerJ*, 3:e1455, 2015.

[23] Morton M Silverman and Alan L Berman. Training for suicide risk assessment and suicide risk formulation. *Academic Psychiatry*, 38:526–537, 2014.

[24] Nighat Z Khan and Muhammad Ali Javed. Use of artificial intelligence-based strategies for assessing suicidal behavior and mental illness: A literature review. *Cureus*, 14(7), 2022.

[25] Hossein Hassani, Emmanuel Sirimal Silva, Stephane Unger, Maedeh TajMazinani, and Stephen Mac Feely. Artificial intelligence (ai) or intelligence augmentation (ia): what is the future? *Ai*, 1(2):8, 2020.

[26] S Haripriya and LC Manikandan. A study on artificial intelligence technologies and its applications. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 6(4):336–344, 2020.

[27] Aisha Ali Hawsawi, Neil Nixon, Emily Stewart, and Elena Nixon. Medical students' perceptions of factors associated with their mental health and psychological well-being. *BJPsych Open*, 9(S1):S51–S51, 2023.

[28] Nelao T Mhata, Vuyokazi Ntlantsana, Andrew M Tomita, Kissah Mwambene, and Shamima Saloojee. Prevalence of depression, anxiety and burnout in medical students at the university of namibia. *South African Journal of Psychiatry*, 29(1), 2023.

[29] Md Ashiqur Rahman Ashiq, Pradip Sen Gupta, Md Abdullah Al Jubayer Biswas, Nowreen Ahmed, Mst Sadia Sultana, Bikona Ghosh, and M Tasdik Hasan. Depression, anxiety, stress, and fear of covid-19 among bangladeshi medical students during the first wave of the pandemic: a mixed-methods study. *Frontiers in Psychiatry*, 14:1142724, 2023.

[30] Muktha Nair, Nathaniel Moss, Amna Bashir, David Garate, Devon Thomas, Shangyi Fu, Daniel Phu, and Christine Pham. Mental health trends among medical students. In *Baylor University Medical Center Proceedings*, volume 36, pages 408–410. Taylor & Francis, 2023.

[31] Asta Medisauskaite, Milou EWM Silkens, and Antonia Rich. A national longitudinal cohort study of factors contributing to uk medical students' mental ill-health symptoms. *General Psychiatry*, 36(2):e101004, 2023.

[32] Che Henry Ngwa, John Njuma Libwea, Marion Keinamma, Frankline Sevidzem Wirsiy, Kah Emmanuel Nji, and Limkile Mpofu. Prevalence, risk factors and management of common mental health disorders in cameroon: a systematic review protocol. *BMJ open*, 13(7):e068139, 2023.

[33] Abdolreza Gilavand, Behnoosh Khoshouie, and Samaneh Mohamadpour. Investigating the prevalence of depression, anxiety, stress, and worry, and its relationship with the academic performance of medical students during the covid-19 pandemic in iran. *Jundishapur J Chronic Dis Care*, 12:e137714, 2023.

[34] Gaurav Urs and HR Vinay. Health anxiety and academic performance among undergraduate medical students: A correlational study. *International Health Trends and Perspectives*, 3(1):106–114, 2023.

[35] Jannathul Firdous, ATM Haque, Anis Lyanna, Auni Nawwarah, Fatin Najihah, Nur Iffah Zakiah, Nur Syahirah, et al. Sleep disruption and its impact on academic performance in medical students: A systematic review. *Horizon Research Publishing*, 2023.

[36] Ramachandran Niranjjan, S Nancy, S Gayathri, and Subramaniam Arulvijayavani. Mental health status and substance abuse among medical students in karaikal, puducherry, india. *Bioinformation*, 20(3):292, 2024.

[37] Absar Nazir, Zeeshan Malik, et al. Stress and coping skills among undergraduate medical students-a cross sectional study. *Pakistan Postgraduate Medical Journal*, 34(02):73–76, 2023.

[38] S Delna, C Pradeep, IJ Nirmal Sujitha, and B Lakshmi Dorai. Stress and strategies in coping stress among the medical students of a south indian city. *MGM Journal of Medical Sciences*, 10(1):18–23, 2023.

[39] Jiang-Yan Song, Ya-Qin Gao, Hui-Ping Li, Jie Gao, Jia-Xin Li, Qing-quan Bi, Shu-Mei Li, and Yun Ye. The latent category of resilience and its relationship with the mental health and coping styles of medical students. In *Frontiers in Education*, volume 8, page 1114352. Frontiers Media SA, 2023.

[40] Jennifer Wang, Reagan E Fitzke, Denise D Tran, Jewel Grell, and Eric R Pedersen. Mental health treatment-seeking behaviors in medical students: A mixed-methods approach. *The Journal of Medicine Access*, 7:27550834221147787, 2023.

[41] Ranna Nash, Brandon Conner, Katelyn Fellows, Brooke Clemmensen, Robert Gullickson, and Sharon Goldrup. Barriers in medical education: A scoping review of common themes for medical students with disabilities. *Discover Education*, 1(1):4, 2022.

[42] Barnabás Oláh, Bence Márk Rádi, and Karolina Kósa. Barriers to seeking mental help and interventions to remove them in medical school during the covid-19 pandemic: perspectives of students. *International Journal of Environmental Research and Public Health*, 19(13):7662, 2022.

[43] Ursula Moffitt, Dalal Katsiaficas, Negin Ghavami, Imani Minor, Dayanara Padilla, and Leoandra Onnie Rogers. Intersectionality and identity: A systematic review and qualitative analysis of us research in psychological science. *Identity*, 23(4):288–313, 2023.

[44] Margarita Alegría and Michelle Cheng. Intersectional approaches are essential to identify the multiple sources of oppression. *Journal of psychopathology and clinical science*, 132(5):590, 2023.

[45] Alberto Sánchez-Navarro, María F. Macías, Eduardo Sánchez, José M. Zepeda, Felix Osuna, and Miriam C. Pérez. Prevalence and risk factors associated with mental health disorders among medical students. *International Journal Of Community Medicine And Public Health*, 9(2):1010–1016, 2022.

[46] Brandon Maser, Marlon Danilewitz, Eva Guérin, Leanne Findlay, and Erica Frank. Medical student psychological distress and mental illness relative to the general population: a canadian cross-sectional survey. *Academic Medicine*, 94(11):1781–1791, 2019.

[47] Daniel Pagnin and Valeria De Queiroz. Comparison of quality of life between medical students and young general populations. *Education for Health*, 28(3):209–212, 2015.

[48] Henry Li, Tushar Upreti, Victor Do, Erica Dance, Melanie Lewis, Ryan Jacobson, and Aviva Goldberg. Measuring wellbeing: A scoping review of metrics and studies measuring medical student wellbeing across multiple timepoints. *Medical Teacher*, 46(1):82–101, 2024.

[49] Ilona Kaisti, Petri Kulmala, Mirka Hintsanen, Tuula Hurtig, Saara Repo, Tiina Paunio, Jouko Miettunen, Anu-Helmi Halt, and Erika Jääskeläinen. The effects of mindfulness-based interventions in medical students: a systematic review. *Advances in Health Sciences Education*, 29(1):245–271, 2024.

[50] Laura D'Adamo, Layna Paraboschi, Anne Claire Grammer, Molly Fennig, Andrea K Graham, Lauren H Yaeger, Michelle G Newman, Denise E Wilfley, C Barr Taylor, Daniel Eisenberg, et al. Reach and uptake of digital mental health interventions based on cognitive-behavioral therapy for college students: a systematic review. *Journal of behavioral and cognitive therapy*, 33(2):97–117, 2023.

[51] Ikeoluwapo Kendra Bolanle-Rufai. Depression and suicidal ideation among medical students: A literature review. *DOKITA*, 40(2):12–16, 2023.

[52] Diyaz Syauki Ikhsan, Bintang Arroyantri Prananjaya, Syarifah Aini, Puji Rizki Suryani, Verdy Cendana, and Nicholas Tze Ping Pang. Suicide in medical student: How to prevent? *JPS*, 11(2):10–20473, 2022.

[53] Fidel Cacheda, Diego Fernandez, Francisco J Novoa, and Victor Carneiro. Early detection of depression: social network analysis and random forest techniques. *Journal of medical Internet research*, 21(6):e12554, 2019.

[54] David Owen, Dimosthenis Antypas, Athanasios Hassoulas, Antonio F Pardiñas, Luis Espinosa-Anke, Jose Camacho Collados, et al. Enabling early health care intervention by detecting depression in users of web-based forums using language models: longitudinal analysis and evaluation. *JMIR AI*, 2(1):e41205, 2023.

[55] Tapotosh Ghosh, Md Hasan Al Banna, Md Jaber Al Nahian, Mohammed Nasir Uddin, M Shamim Kaiser, and Mufti Mahmud. An attention-based hybrid architecture with explainability for depressive social media text detection in bangla. *Expert Systems with Applications*, 213:119007, 2023.

[56] Hui Zhang, Hong Wang, Shu Han, Wei Li, and Luhe Zhuang. Detecting depression tendency with multimodal features. *Computer Methods and Programs in Biomedicine*, 240:107702, 2023.

[57] Ju Chun Cheng and Arbee LP Chen. Multimodal time-aware attention networks for depression detection. *Journal of Intelligent Information Systems*, 59(2):319–339, 2022.

[58] Lu Ren, Hongfei Lin, Bo Xu, Shaowu Zhang, Liang Yang, and Shichang Sun. Depression detection on reddit with an emotion-based attention network: algorithm development and validation. *JMIR medical informatics*, 9(7):e28754, 2021.

[59] Xiaofeng Wang, Shuai Chen, Tao Li, Wanting Li, Yejie Zhou, Jie Zheng, Qingcai Chen, Jun Yan, Buzhou Tang, et al. Depression risk prediction for chinese microblogs via deep-learning methods: Content analysis. *JMIR medical informatics*, 8(7):e17958, 2020.

[60] Ah Young Kim, Eun Hye Jang, Seung-Hwan Lee, Kwang-Yeon Choi, Jeon Gue Park, and Hyun-Chool Shin. Automatic depression detection using smartphone-based text-dependent speech signals: deep convolutional neural network approach. *Journal of medical Internet research*, 25:e34474, 2023.

[61] Alfredo Daza Vergaray, Juan Carlos Herrera Miranda, Juana Bobadilla Cornelio, Atilio Rubén López Carranza, and Carlos Fidel Ponce Sánchez. Predicting the depression in university students using stacking ensemble techniques over oversampling method. *Informatics in medicine unlocked*, 41:101295, 2023.

[62] Akshi Kumar, Saurabh Raj Sangwan, Anshika Arora, and Varun G Menon. Depress-dcnf: A deep convolutional neuro-fuzzy model for detection of depression episodes using iomt. *Applied Soft Computing*, 122:108863, 2022.

[63] Amal Bouzid, Abdulrahman Almidani, Maria Zubrikhina, Altyngul Kamzanova, Burcu Yener Ilce, Manzura Zholdassova, Ayesha M Yusuf, Poorna Manasa Bhamidimarri, Hamid A AlHaj, Almira Kustubayeva, et al. Integrative bioinformatics and artificial intelligence analyses of transcriptomics data identified genes associated with major depressive disorders including nrg1. *Neurobiology of stress*, 26:100555, 2023.

[64] Daniela Caldirola, Silvia Daccò, Francesco Cuniberti, Massimiliano Grassi, Alessandra Alciati, Tatiana Torti, and Giampaolo Perna. First-onset major depression during the covid-19 pandemic: A predictive machine learning model. *Journal of Affective Disorders*, 310:75–86, 2022.

[65] Kyoung-Sae Na, Seo-Eun Cho, Zong Woo Geem, and Yong-Ku Kim. Predicting future onset of depression among community dwelling adults in the republic of korea using a machine learning algorithm. *Neuroscience Letters*, 721:134804, 2020.

[66] Nikhil Marriwala, Deepti Chaudhary, et al. A hybrid model for depression detection using deep learning. *Measurement: Sensors*, 25:100587, 2023.

[67] Kirill Milintsevich, Kairit Sirts, and Gaël Dias. Towards automatic text-based estimation of depression through symptom prediction. *Brain informatics*, 10(1):4, 2023.

[68] Kennedy Opoku Asare, Isaac Moshe, Yannik Terhorst, Julio Vega, Simo Hosio, Harald Baumeister, Laura Pulkki-Råback, and Denzil Ferreira. Mood ratings and digital biomarkers from smartphone and wearable data differentiates and predicts depression status: A longitudinal data analysis. *Pervasive and Mobile Computing*, 83:101621, 2022.

[69] Ada Dogrucu, Alex Perucic, Anabella Isaro, Damon Ball, Ermal Toto, Elke A Rundensteiner, Emmanuel Agu, Rachel Davis-Martin, and Edwin Boudreaux. Moodable: On feasibility of instantaneous depression assessment using machine learning on voice samples with retrospectively harvested smartphone and social media data. *Smart Health*, 17:100118, 2020.

[70] Kennedy Opoku Asare, Yannik Terhorst, Julio Vega, Ella Peltonen, Eemil Lagerspetz, and Denzil Ferreira. Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study. *JMIR mHealth and uHealth*, 9(7):e26540, 2021.

[71] Laura M Lyall, Natasha Sangha, Xingxing Zhu, Donald M Lyall, Joey Ward, Rona J Strawbridge, Breda Cullen, and Daniel J Smith. Subjective and objective sleep and circadian parameters as predictors of depression-related outcomes: a machine learning approach in uk biobank. *Journal of Affective Disorders*, 335:83–94, 2023.

[72] Chenyang Zhang, Xiaofei Chen, Song Wang, Junjun Hu, Chunpeng Wang, and Xin Liu. Using catboost algorithm to identify middle-aged and elderly depression, national health and nutrition examination survey 2011–2018. *Psychiatry research*, 306:114261, 2021.

[73] Rokeya Siddiqua, Nusrat Islam, Jarba Farnaz Bolaka, Riasat Khan, and Sifat Momen. Aida: Artificial intelligence based depression assessment applied to bangladeshi students. *Array*, 18:100291, 2023.

[74] Lvchun Cui, Chenglei Wang, Zhiguo Wu, Daihui Peng, Jingjing Huang, Chen Zhang, Jia Huang, Wu Hong, Yong Wang, Jun Chen, et al. Symptomatology differences of major depression in psychiatric versus general hospitals: A machine learning approach. *Journal of affective disorders*, 260:349–360, 2020.

[75] Yu Wu, Binbin Su, Chen Chen, Yihao Zhao, Panliang Zhong, and Xiaoying Zheng. Urban-rural disparities in the prevalence and trends of depressive symptoms among chinese elderly and their associated factors. *Journal of Affective Disorders*, 340:258–268, 2023.

[76] Md Zia Uddin, Kim Kristoffer Dysthe, Asbjørn Følstad, and Petter Bae Brandtzaeg. Deep learning for prediction of depressive symptoms in a large textual dataset. *Neural Computing and Applications*, 34(1):721–744, 2022.

[77] Kuk-In Jang, Sungkean Kim, Jeong-Ho Chae, and Chany Lee. Machine learning-based classification using electroencephalographic multi-paradigms between drug-naïve patients with depression and healthy controls. *Journal of Affective Disorders*, 338:270–277, 2023.

[78] Danniel Shazmeer Bin Abdul Hamid, SB Goyal, and Pradeep Bedi. Integration of deep learning for improved diagnosis of depression using eeg and facial features. *Materials Today: Proceedings*, 80:1965–1969, 2023.

[79] Ayan Seal, Rishabh Bajpai, Mohan Karnati, Jagriti Agnihotri, Anis Yazidi, Enrique Herrera-Viedma, and Ondrej Krejcar. Benchmarks for machine learning in depression discrimination using electroencephalography signals. *Applied Intelligence*, 53(10):12666–12683, 2023.

[80] Sukit Suparatpinyo and Nuanwan Soonthornphisaj. Smart voice recognition based on deep learning for depression diagnosis. *Artificial Life and Robotics*, 28(2):332–342, 2023.

[81] Heejung Kim, SungHee Lee, SangEun Lee, Soyun Hong, HeeJae Kang, Namhee Kim, et al. Depression prediction by using ecological momentary assessment, actiwatch data, and machine learning: observational study on older adults living alone. *JMIR mHealth and uHealth*, 7(10):e14149, 2019.

[82] Jae Sung Kim, Bohyun Wang, Meelim Kim, Jung Lee, Hyungjun Kim, Danyeul Roh, Kyung Hwa Lee, Soon-Beom Hong, Joon Shik Lim, Jae-Won Kim, et al. Prediction of diagnosis and treatment response in adolescents with depression by using a smartphone app and deep learning approaches: usability study. *JMIR formative research*, 7(1):e45991, 2023.

[83] Mudasir Ahmad Wani, Mohammad A ELAffendi, Kashish Ara Shakil, Ali Shariq Imran, and Ahmed A Abd El-Latif. Depression screening in humans with ai and deep learning techniques. *IEEE transactions on computational social systems*, 10(4):2074–2089, 2022.

[84] Duyan Geng, Qiang An, Zhigang Fu, Chao Wang, and Hongxia An. Identification of major depression patients using machine learning models based on heart rate variability during sleep stages for pre-hospital screening. *Computers in Biology and Medicine*, 162:107060, 2023.

[85] Yuri Rykov, Thuan-Quoc Thach, Iva Bojic, George Christopoulos, and Josip Car. Digital biomarkers for depression screening with wearable devices: cross-sectional study with machine learning modeling. *JMIR mHealth and uHealth*, 9(10):e24872, 2021.

[86] Tahsin Mullick, Ana Radovic, Sam Shaaban, Afsaneh Doryab, et al. Predicting depression in adolescents using mobile and wearable sensors: multimodal machine learning–based exploratory study. *JMIR Formative Research*, 6(6):e35807, 2022.

[87] Yang S Liu, Stefani Chokka, Bo Cao, and Pratap R Chokka. Screening for bipolar disorder in a tertiary mental health centre using earlydetect: A machine learning-based pilot study. *Journal of Affective Disorders Reports*, 6:100215, 2021.

[88] Ran Bai, Le Xiao, Yu Guo, Xuequan Zhu, Nanxi Li, Yashen Wang, Qinqin Chen, Lei Feng, Yinghua Wang, Xiangyi Yu, et al. Tracking and monitoring mood stability of patients with major depressive disorder by machine learning models using passive digital data: prospective naturalistic multicenter study. *JMIR mHealth and uHealth*, 9(3):e24365, 2021.

[89] Zhiwei Bao, Xinyi Zhao, Jingjing Li, Guanghua Zhang, Hairong Wu, Yuping Ning, Ming D Li, and Zhongli Yang. Prediction of repeated-dose intravenous ketamine response in major depressive disorder using the gwas-based machine learning approach. *Journal of Psychiatric Research*, 138:284–290, 2021.

[90] Aidan Cousins, Lucas Nakano, Emma Schofield, and Rasa Kabaila. A neural network approach to optimising treatments for depression using data from specialist and community psychiatric services in australia, new zealand and japan. *Neural Computing and Applications*, 35(16):11497–11516, 2023.

[91] Muhammad Khubayeeb Kabir, Maisha Islam, Anika Nahian Binte Kabir, Adiba Haque, and Md Khalilur Rhaman. Detection of depression severity using bengali social media posts on mental health: study using natural language processing techniques. *JMIR Formative Research*, 6(9):e36118, 2022.

[92] Mariko Makhmutova, Raghu Kainkaryam, Marta Ferreira, Jae Min, Martin Jaggi, Ieuan Clay, et al. Predicting changes in depression severity using the psyche-d (prediction of severity change-depression) model involving person-generated health data: longitudinal case-control observational study. *JMIR mHealth and uHealth*, 10(3):e34148, 2022.

[93] Soumitra Ghosh, Asif Ekbal, and Pushpak Bhattacharyya. A multitask framework to detect depression, sentiment and multi-label emotion from suicide notes. *Cognitive Computation*, 14(1):110–129, 2022.

[94] Zachary A Cordner, Dean F MacKinnon, and J Raymond DePaulo Jr. The care of patients with complex mood disorders. *Focus*, 18(2):129–138, 2020.

[95] Koushik Sinha Deb. Healing the healer–a 10-year journey of supporting a medical student with bipolar disorder. *Indian Journal of Social Psychiatry*, 36(1):5–10, 2020.

[96] Ingrid Vargas-Huicochea, Rebeca Robles-García, Carlos Berlanga, Carlos-Alfonso Tovilla-Zarate, Nicolás Martínez-López, and Ana Fresán. Mental health literacy about bipolar disorder and schizophrenia among medical students: A comparative study of illness recognition, treatment, and attitudes according to perception of aggressiveness-dangerousness. *Salud mental*, 40(4):141–148, 2017.

[97] Casey C Bennett, Mindy K Ross, EuGene Baek, Dohyeon Kim, and Alex D Leow. Predicting clinically relevant changes in bipolar disorder outside the clinic walls based on pervasive

technology interactions via smartphone typing dynamics. *Pervasive and Mobile Computing*, 83:101598, 2022.

[98] Yang Liu, Jeffrey Hankey, Bo Cao, and Pratap Chokka. Screening for major depressive disorder in a tertiary mental health centre using earlydetect: A machine learning-based pilot study. *Journal of affective disorders reports*, 3:100062, 2021.

[99] Jonas Busk, Maria Faurholt-Jepsen, Mads Frost, Jakob E Bardram, Lars Vedel Kessing, and Ole Winther. Forecasting mood in bipolar disorder from smartphone self-assessments: hierarchical bayesian approach. *JMIR mHealth and uHealth*, 8(4):e15028, 2020.

[100] Luisa de Siqueira Rotenberg, Renato Gomes Borges-Junior, Beny Lafer, Rogerio Salvini, and Rodrigo da Silva Dias. Exploring machine learning to predict depressive relapses of bipolar disorder patients. *Journal of Affective Disorders*, 295:681–687, 2021.

[101] Kathryn L Lovero, Palmira F Dos Santos, Amalio X Come, Milton L Wainberg, and Maria A Oquendo. Suicide in global mental health. *Current psychiatry reports*, 25(6):255–262, 2023.

[102] Dheeraj Kodati and Ramakrishnudu Tene. Identifying suicidal emotions on social media through transformer-based deep learning. *Applied Intelligence*, 53(10):11885–11917, 2023.

[103] Shaoxiong Ji, Xue Li, Zi Huang, and Erik Cambria. Suicidal ideation and mental disorder detection with attentive relation networks. *Neural Computing and Applications*, 34(13):10309–10319, 2022.

[104] En-Liang Wu, Chia-Yi Wu, Ming-Been Lee, Kuo-Chung Chu, and Ming-Shih Huang. Development of internet suicide message identification and the monitoring-tracking-rescuing model in taiwan. *Journal of affective disorders*, 320:37–41, 2023.

[105] Jihad S Obeid, Jennifer Dahne, Sean Christensen, Samuel Howard, Tami Crawford, Lewis J Frey, Tracy Stecker, and Brian E Bunnell. Identifying and predicting intentional self-harm in electronic health record clinical notes: deep learning approach. *JMIR medical informatics*, 8(7):e17784, 2020.

[106] Hannah Yao, Sina Rashidian, Xinyu Dong, Hongyi Duanmu, Richard N Rosenthal, and Fusheng Wang. Detection of suicidality among opioid users on reddit: machine learning–based approach. *Journal of medical internet research*, 22(11):e15293, 2020.

[107] Syed Tanzeel Rabani, Akib Mohi Ud Din Khanday, Qamar Rayees Khan, Umar Ayoub Hajam, Ali Shariq Imran, and Zenun Kastrati. Detecting suicidality on social media: Machine learning at rescue. *Egyptian Informatics Journal*, 24(2):291–302, 2023.

[108] Hayoung Kim Donnelly, Yoonsun Han, Suna Kim, and Dong Hun Lee. Predictors of suicide ideation among south korean adolescents: A machine learning approach. *Journal of affective disorders*, 329:557–565, 2023.

[109] Juliet Beni Edgcomb, Chi-hong Tseng, Mengtong Pan, Alexandra Klomhaus, and Bonnie T Zima. Assessing detection of children with suicide-related emergencies: evaluation and development of computable phenotyping approaches. *JMIR mental health*, 10:e47084, 2023.

[110] Anmol Arora, Louis Bojko, Santosh Kumar, Joseph Lillington, Sukhmeet Panesar, and Bruno Petrungaro. Assessment of machine learning algorithms in national data to classify the risk of self-harm among young adults in hospital: a retrospective study. *International Journal of Medical Informatics*, 177:105164, 2023.

[111] Marika Cusick, Prakash Adekkanattu, Thomas R Campion Jr, Evan T Sholle, Annie Myers, Samprit Banerjee, George Alexopoulos, Yanshan Wang, and Jyotishman Pathak. Using weak supervision and deep learning to classify clinical notes for identification of current suicidal ideation. *Journal of psychiatric research*, 136:95–102, 2021.

[112] Polona Rus Prelog, Teodora Matić, Peter Pregelj, and Aleksander Sadikov. A pilot predictive model based on covid-19 data to assess suicidal ideation indirectly. *Journal of Psychiatric Research*, 163:318–324, 2023.

[113] Kelly L Zuromski, Samantha L Bernecker, Carol Chu, Chelsey R Wilks, Peter M Gutierrez, Thomas E Joiner, Howard Liu, James A Naifeh, Matthew K Nock, Nancy A Sampson, et al. Pre-deployment predictors of suicide attempt during and after combat deployment: Results from the army study to assess risk and resilience in servicemembers. *Journal of psychiatric research*, 121:214–221, 2020.

[114] Cecilia Lao, Jo Lane, Hanna Suominen, et al. Analyzing suicide risk from linguistic features in social media: evaluation study. *JMIR formative research*, 6(8):e35563, 2022.

[115] Adonias C de Oliveira, Evandro JS Diniz, Silmar Teixeira, and Ariel S Teles. How can machine learning identify suicidal ideation from user's texts? towards the explanation of the boamente system. *Procedia Computer Science*, 206:141–150, 2022.

[116] Hongxia Lu, Alex Barrett, Albert Pierce, Jianwei Zheng, Yun Wang, Chun Chiang, and Cyril Rakovski. Predicting suicidal and self-injurious events in a correctional setting using ai algorithms on unstructured medical notes and structured data. *Journal of psychiatric research*, 160:19–27, 2023.

[117] Seo-Eun Cho, Zong Woo Geem, and Kyoung-Sae Na. Prediction of suicide among 372,813 individuals under medical check-up. *Journal of psychiatric research*, 131:9–14, 2020.

[118] Juliet Beni Edgcomb, Trevor Shaddox, Gerhard Hellemann, and John O Brooks III. Predicting suicidal behavior and self-harm after general hospitalization of adults with serious mental illness. *Journal of psychiatric research*, 136:515–521, 2021.

[119] Junren Wang, Jiajun Qiu, Ting Zhu, Yu Zeng, Huazhen Yang, Yanan Shang, Jin Yin, Yajing Sun, Yuanyuan Qu, Unnur A Valdimarsdóttir, et al. Prediction of suicidal behaviors in the middle-aged population: machine learning analyses of uk biobank. *JMIR public health and surveillance*, 9(1):e43419, 2023.

[120] Raymond Su, James Rufus John, and Ping-I Lin. Machine learning-based prediction for self-harm and suicide attempts in adolescents. *Psychiatry research*, 328:115446, 2023.

[121] Sooyeon Min, Daun Shin, Sang Jin Rhee, C Hyung Keun Park, Jeong Hun Yang, Yoojin Song, Min Ji Kim, Kyungdo Kim, Won Ik Cho, Oh Chul Kwon, et al. Acoustic analysis of speech for screening for suicide risk: machine learning classifiers for between-and within-person evaluation of suicidality. *Journal of medical internet research*, 25:e45456, 2023.

[122] Brian Stasak, Julien Epps, Heather T Schatten, Ivan W Miller, Emily Mower Provost, and Michael F Armey. Read speech voice quality and disfluency in individuals with recent suicidal ideation or suicide attempt. *Speech Communication*, 132:10–20, 2021.

[123] Wesllei Felipe Heckler, Luan Paris Feijó, Juliano Varella de Carvalho, and Jorge Luis Victória Barbosa. Thoth: An intelligent model for assisting individuals with suicidal ideation. *Expert Systems with Applications*, 233:120918, 2023.

[124] Milena Ilic and Irena Ilic. Worldwide suicide mortality trends (2000-2019): A joinpoint regression analysis. *World journal of psychiatry*, 12(8):1044, 2022.

[125] Mohammed A Mamun, Jannatul Mawa Misti, and Mark D Griffiths. Suicide of bangladeshi medical students: Risk factor trends based on bangladeshi press reports. *Asian Journal of Psychiatry*, 48:101905, 2020.

[126] Carol Swetlik and Kathleen N Franco. Medical student suicide: an assessment of risk factors and prevention strategies. In *Suicide Prevention: A Practical Guide for the Practitioner*, pages 251–263. Springer, 2018.

[127] Chanhee Seo, Christina Di Carlo, Selina Xiangxu Dong, Karine Fournier, and Kay-Anne Haykal. Risk factors for suicidal ideation and suicide attempt among medical students: A meta-analysis. *PloS one*, 16(12):e0261785, 2021.

[128] Lisa S Rotenstein, Marco A Ramos, Matthew Torre, J Bradley Segal, Michael J Peluso, Constance Guille, Srijan Sen, and Douglas A Mata. Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: a systematic review and meta-analysis. *Jama*, 316(21):2214–2236, 2016.

[129] Wanjie Tang, Yukun Kang, Jiuping Xu, and Tao Li. Associations of suicidality with adverse life events, psychological distress and somatic complaints in a chinese medical student sample. *Community mental health journal*, 56:635–644, 2020.

[130] Azam Naghavi, Tobias Teismann, Zahra Asgari, Mohammad Reza Mohebbian, Marjan Mansourian, and Miguel Ángel Mañanas. Accurate diagnosis of suicide ideation/behavior using robust ensemble machine learning: a university student population in the middle east and north africa (mena) region. *Diagnostics*, 10(11):956, 2020.

[131] Bora Kim, Younghoon Kim, C Hyung Keun Park, Sang Jin Rhee, Young Shin Kim, Bennett L Leventhal, Yong Min Ahn, Hyojung Paik, et al. Identifying the medical lethality of suicide attempts using network analysis and deep learning: nationwide study. *JMIR medical informatics*, 8(7):e14500, 2020.

[132] Sultan Mahmud, Md Mohsin, Abdul Muyeed, Shaila Nazneen, Md Abu Sayed, Nabil Murshed, Tajrin Tahrin Tonmon, and Ariful Islam. Machine learning approaches for predicting suicidal behaviors among university students in bangladesh during the covid-19 pandemic: A cross-sectional study. *Medicine*, 102(28):e34285, 2023.

[133] Yanmei Shen, Wenyu Zhang, Bella Siu Man Chan, Yaru Zhang, Fanchao Meng, Elizabeth A Kennon, Hanjing Emily Wu, Xuerong Luo, and Xiangyang Zhang. Detecting risk of suicide attempts among chinese medical college students using a machine learning algorithm. *Journal of affective disorders*, 273:18–23, 2020.

[134] Md. Omar Faruk, Tazkia Zaman Shova, Ismail Hosen, Firoj al Mamun, M. Rasheduzzaman, and Mohammed Mamun. Dataset on mental health, suicidal behaviors, literacy and stigma in bangladeshi dental students, 2021.

[135] Rifat Jahan Chomon. Depression and suicidal ideation among medical students in a private medical college of bangladesh. a cross sectional web based survey. *PloS one*, 17(4):e0265367, 2022.

[136] Suwatthanachai Phomprasith, Nuntaporn Karawekpanyawong, Kanokporn Pinyopornpanish, Wichuda Jiraporncharoen, Benchalak Maneeton, Phichayut Phinyo, and Suppachai Lawanaskol. Prevalence and associated factors of depression in medical students in a northern thailand university: a cross-sectional study. In *Healthcare*, volume 10, page 488. MDPI, 2022.

[137] Gustavo Sosa-Cabrera, Santiago Gómez-Guerrero, Miguel García-Torres, and Christian E Schaerer. Feature selection: A perspective on inter-attribute cooperation. *International Journal of Data Science and Analytics*, 17(2):139–151, 2024.

[138] Sébastien Déjean, Radu Tudor Ionescu, Josiane Mothe, and Md Zia Ullah. Forward and backward feature selection for query performance prediction. In *Proceedings of the 35th annual ACM symposium on applied computing*, pages 690–697, 2020.

[139] Giorgos Borboudakis and Ioannis Tsamardinos. Forward-backward selection with early dropping. *Journal of Machine Learning Research*, 20(8):1–39, 2019.

[140] Khondaker Abdullah-Al-Mamun. *Pattern identification of movement related states in biosignals*. PhD thesis, University of Southampton, 2013.

[141] Mariachiara Mecati, Marco Torchiano, Antonio Vetrò, and Juan Carlos De Martin. Measuring imbalance on intersectional protected attributes and on target variable to forecast unfair classifications. *IEEE Access*, 11:26996–27011, 2023.

[142] Alberto Fernández, Salvador Garcia, Francisco Herrera, and Nitesh V Chawla. Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of artificial intelligence research*, 61:863–905, 2018.

[143] Manuel Rausch and Michael Zehetleitner. Should metacognition be measured by logistic regression? *Consciousness and cognition*, 49:291–312, 2017.

[144] Hayoung Kim Donnelly, Yoonsun Han, Suna Kim, and Dong Hun Lee. Predictors of suicide ideation among south korean adolescents: A machine learning approach. *Journal of affective disorders*, 329:557–565, 2023.

[145] Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.

[146] Nico Potyka, Xiang Yin, and Francesca Toni. Explaining random forests using bipolar argumentation and markov networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 9453–9460, 2023.

[147] Andreas Mueller, Gian Candrian, Juri D Kropotov, Valery A Ponomarev, and Gian-Marco Baschera. Classification of adhd patients on the basis of independent erp components using

a machine learning system. In *Nonlinear biomedical physics*, volume 4, pages 1–12. Springer, 2010.

[148] Dongsheng Che, Qi Liu, Khaled Rasheed, and Xiuping Tao. Decision tree and ensemble learning algorithms with their applications in bioinformatics. *Software tools and algorithms for biological systems*, pages 191–199, 2011.

[149] Sayali D Jadhav and HP Channe. Comparative study of k-nn, naive bayes and decision tree classification techniques. *International Journal of Science and Research (IJSR)*, 5(1):1842–1845, 2016.

[150] Juliet Beni Edgcomb, Trevor Shaddox, Gerhard Hellemann, and John O Brooks III. Predicting suicidal behavior and self-harm after general hospitalization of adults with serious mental illness. *Journal of psychiatric research*, 136:515–521, 2021.