LEVERAGING MACHINE LEARNING FOR PREDICTIVE ANALYSIS OF DRUG ADDICTION IN BANGLADESH

BY

Fahmida Akter
ID: 221-17-517

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Master of Management Information System

Supervised By

Abdus Sattar
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2025

APPROVAL

This Project titled **"Leveraging Machine Learning for Predictive Analysis of Drug Addiction in Bangladesh"**, submitted by **Fahmida Akter**, ID No: **221-17-517** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of MS in Management Information System and approved as to its style and contents. The presentation has been held on 11 January Saturday, 2025.

BOARD OF EXAMINERS

Dr. S. M. Aminul Haque — **Chairman**
Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. Naznin Sultana — **Internal Examiner**
Associate Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mr. Md. Sadekur Rahman — **Internal Examiner**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mr. Nazibur Rahman — **External Examiner**
Technical Lead - Database Administrator,
Wipro Bangladesh Telenor - Grameen Phone

DECLARATION

I hereby declare that, this project has been done by me under the supervision of **Abdus Sattar, Assistant Professor, Department of CSE,** Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Abdus Sattar
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Fahmida Akter

Fahmida Akter
ID: 221-17-517
Department of MIS
Daffodil International University

ACKNOWLEDGEMENT

First I express my heartiest thanks and gratefulness to Almighty Allah for His divine blessing which makes me possible to complete the final year project/internship successfully.

I really grateful and wish my profound indebtedness to **Abdus Sattar, Assistant Professor**, Department of CSE, Daffodil International University, Dhaka, deep knowledge & keen interest of my supervisor in the field of Machine Learning to carry out this project. His endless patience, scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

I would like to express my heartiest gratitude to **Dr.Sheak Rashed Haider Noori**, **Head,** Department of CSE, for his kind help to finish our project and also to other faculty members and the staffs of MIS department of Daffodil International University.

Finally, I must acknowledge with due respect the constant support and patients of my parents.

ABSTRACT

Drug addiction has become a critical health concern, especially among the youth of Bangladesh. According to AsiaNews, there are more than 8 million drug addicts, out of which a large portion consists of young people. Due to this growing menace, some effective preventive steps need to be taken. This research examines some important factors that differentiate between addicted and non-addicted people to help in targeted interventions. This study is based on 1,624 individual responses, collected through an online survey containing a wide array of demographic and behavioral attributes. The dataset includes participants from Dhaka and Sylhet, aged 15 to 27 years, predominantly students. In this paper, a machine learning-based approach was followed for analyzing and predicting the risk of addiction. Models used in this work are Multilayer Perceptron (MLP), Extreme Learning Machine (ELM), CatBoost, and standard classifiers. Ensemble techniques were also adopted, namely Blending (Neural Networks, Gradient Boosting, and K-Nearest Neighbors) and Voting (Logistic Regression, Random Forest, Support Vector Machine). Models were evaluated as well, and their performances corresponded to 98.15% for Blending and CatBoost, 93.85% for Voting, and 99.38% for both MLP and ELM. Top-working models will help identify people in vulnerable positions toward addiction, evaluate the state of health and psychological condition of a person, which will be an indicator for acting in prevention. Outcome variables such as these from the current study are critical in deducing further risks for addiction, especially in targeted public health policies and intervention programs. With a special emphasis on current states of mental and physical health, the high performing models, especially MLP and ELM, proved as reliable tools for assessing risk for addiction. It outlines the role of machine learning in the transformation that views and methods toward the drug addiction crisis among Bangladeshi youth have been causing, with a data-driven basis for prevention and rehabilitation efforts.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES	PAGE NO

LIST OF TABLES

TABLES **PAGE NO**

CHAPTER 1
INTRODUCTION

1.1 Introduction

Addiction is not just a word but, as such, is being an important social problem seriously effecting youths, even elders, bringing hindrance for the entire nation. This is a global challenge, causing severe damage in one's mental and physical health. It also holds strong relations with social as well as family aspects. Most addicts are unable to appreciate how or why they became addicted to drugs. Many drug users take the first dose out of curiosity, under peer pressure, or some mislaid passion. However, addiction is a complex illness, and therefore recovery is equally not easy. More than 80% of drug addicts in Bangladesh include youths ranging from 15 to 35 years. In fact, the causes that can be attributed to first-time drug use are personal dissatisfaction, heartbreak, family conflicts, political instability, and lack of meaningful friendships.

This trend is really alarming and immediate attention is required if this country wants to reduce the impact on its future generations. Avoiding drugs can significantly reduce the likelihood of addiction and help curb its rising prevalence. Drug abuse has become a pervasive issue among young people, stealthily disrupting their lives and routines. The impact of drug addiction extends beyond individuals, contributing to criminal activities and economic strain in Bangladesh.

In 2021, approximately half of those arrested for drug-related crimes were under 25 years old. In addition, it costs the economy about $2 billion yearly due to addiction. The National Drug Control Commission estimates about 2.5 million users in the country, out of which children and teenagers comprise 80%. Substance abuse leads to numerous negative consequences, including health problems, economic decline, and disintegration of the social framework. The earlier the stage at which addiction is addressed, the better, because this may reduce its progress as well as decrease its possibility of recurrence. This research represents the first attempt to apply machine learning in order to study drug addiction in Bangladesh.

It will investigate patterns of addiction within the unique socio-cultural context of the nation and identify the key factors contributing to vulnerability to addiction. The development of machine learning models specific to predicting the risk of addiction in this study aims at improving predictability with a view to targeted interventions. The findings will help in health decision-making by providing data-driven insights into the optimization of resource allocation and treatment strategies. This research points to the potential of technology in transforming addiction analysis and intervention efforts and opens up new avenues for innovative solutions to this pressing challenge.

1.2 Motivation

The high rate of drug addiction among the youth in Bangladesh, as well as its consequences on public health and social stability, are major motivating factors for this research. Recent reports show that more than 80% of drug addicts belong to the age group between 15 and 35 years, with grave consequences for their mental and physical health. This is further compounded by the fact that it contributes to increasing crime rates and economic losses estimated at $2 billion annually. Therefore, the predictors of addiction and the development of specific intervention strategies are very vital in trying to stem this crisis. The research, therefore, leveraging advanced machine learning models, will bridge the gap in existing literature and provide a way in which data-driven insights can guide effective healthcare policies and preventive measures. This should be an attempt at reduction in the socioeconomic cost of drug addiction to give a more healthy and secured future to Bangladesh.

1.3 Rationale of the Study

This research work will hence provide evidence-based insights into the factors that are causing drug addiction among Bangladeshi youth, focusing on the identification of predictors and high-risk behaviors. The research will make use of advanced machine learning techniques to analyze socio-demographic and behavioral data of affected individuals to find out the pattern and trend. Such knowledge is important in the sense that it raises awareness about the need for targeted interventions to try and stem this growing public health crisis. The ultimate goal is to provide policymakers, healthcare providers, and community leaders with actionable data to design and execute strategies that reduce addiction rates, promote mental and physical well-being, and foster a healthier and more resilient society.

1.4 Research Questions

- RQ1: What are the demographic and behavioral factors that distinguish between addicted and non-addicted individuals in Bangladesh?
- RQ2: To what extent can machine learning algorithms such as Multilayer Perceptron (MLP), Extreme Learning Machine (ELM), and CatBoost estimate the probability of drug addiction?
- RQ3: What is the relative performance of the ensemble models such as Blending (Neural Networks + Gradient Boosting + KNN) and Voting (Logistic Regression, Random Forest, SVM) in classifying addiction tendencies?
- RQ4: Can predictive models identify age-specific addiction patterns, especially among the youth aged 15–27 years, in the dataset?
- RQ5: How do the predictive results of this study help in developing targeted interventions to reduce drug addiction among Bangladeshi youth?

1.5 Expected Output

- Development of a robust MLP, CatBoost, and ELM model for accurate identify of drug addiction from collected samples.
- The ensemble methods at an advanced level-Blending that includes Neural Networks, Gradient Boosting, and KNN, and Voting that includes Logistic Regression, Random Forest, and SVM-were strictly compared against standalone models.

- A structured framework for integrating predictive models into public health strategies was thus developed to enable early detection and intervene selectively in cases that run the risk of addiction.
- It points to the identification of critical predictors, such as family dynamics, peer influence, and mental health conditions that could contribute toward designing effective preventive measures and rehabilitation programs among Bangladeshi youth.
- The study lays a great groundwork for further research in integrated advanced machine learning algorithms and varied datasets that could help in understanding the pattern of addiction.

1.6 Project Management and Finance

The research work doesn't get fund from any individuals or organization.

1.7 Report Layout

In Chapter 1, the introduction, objectives, and key research inquiries of the study are outlined. In Chapter 2, concise synopses of the literature review are provided. In Chapter 3, the proposed methodology is described in detail. In Chapter 4, the experimental outcomes of the paper are described and examined. The fifth chapter discusses the sustainability plan, societal and environmental impacts, and ethical considerations. The sixth chapter concludes the present investigation and outlines a strategy for subsequent endeavors.

CHAPTER 2
BACKGROUND

2.1 Preliminaries/Terminologies

Recent machine learning and data analytics developments have opened new avenues for finding fresh approaches to the problem of drug addiction among youth. By using predictive analytics, researchers have come up with advanced systems that can identify and classify addiction tendencies based on socio-demographic and behavioral attributes. Such systems are bound to predict risks for addiction quite accurately with the integration of various machine learning models, including MLP and ELM, among others, along with ensemble techniques, considering peer influence, mental health, and family dynamics. These developments reflect a transformational leap in the field of addiction research that will be providing actionable insight into potentially transforming public health policies and interventions to meet this evolving challenge with a view toward eradication.

2.2 Related works

Various machine learning models used different algorithms and features for the prediction of AUD. Kinreich et al., [2019], employed the Support Vector Machine classification for symptoms of AUD based on multidimensional factors such as gender, ethnicity, and age and realized remarkable accuracy in African American populations. Mary et al. [2019], utilized a decision tree classifier to classify individuals as either therapy seekers or nontreatment seekers, and the Alternating Decision Tree framework emerged with the best performance. Using different measures, Lee et al. [2019], developed a decision tree model that predicted AUD treatment-seeking behavior. Guggenmos et al. [2020], enhanced the diagnostic classification of alcohol dependence by integrating neuroimaging data. Other SUDs have also been vastly explored using different datasets and algorithms for their prediction with the help of machine learning.Various machine learning models used different algorithms and features for the prediction of AUD. Kinreich et al., [2019], employed the Support Vector Machine classification for symptoms of AUD based on multidimensional factors such as gender, ethnicity, and age and realized remarkable accuracy in African American populations. Mary et al. [2019], utilized a decision tree classifier to classify individuals as either therapy seekers or nontreatment seekers, and the Alternating Decision Tree framework emerged with the best performance. Using different measures, Lee et al. [2019], developed a decision tree model that predicted AUD treatment-seeking behavior. Guggenmos et al. [2020], enhanced the diagnostic classification of alcohol dependence by integrating neuroimaging data. Other SUDs have also been vastly explored using different datasets and algorithms for their prediction with the help of machine learning.Kumari et al. [2018], developed a model with two ANN modules for predicting alcohol consumption, attaining notable accuracy. Jing et al. [2020], used the random forest algorithm in order to predict SUD among participants aged 10-22 years, highlighting critical predictors such as psychological dysregulation and health behavior. Beattie & Nicholson, [2021], applied Random Forests on imbalanced heroin use data and reported that the most important predictors were cocaine and crack usage. Similarly, Marcon et al., [2021], analyzed drinking patterns using Elastic Net Regularization, Random Forests, and Neural Networks among medical students in Brazil. Other works have used more specialized methods to enhance these predictions of substance use. For example, Frank et al. [2018], applied logistic regression for the prediction of smoking status based on the results of blood tests; this was the optimal approach in their study.Zhang et al. [2019], proposed the XGBoost algorithm combined with a feature extraction module to enhance daily smoking behavior prediction. Nasir et al. [2021], used integrated data sets of treatment discharge records and crime data, developing random forests, neural networks, and extreme gradient boosting for the precise prediction of treatment completion outcomes. The CNN models used by Mukhtar et al. [2021], on raw EEG data for detecting substance use recorded an astounding 98% accuracy. Further innovative techniques have enriched predictive models for SUDs. Dou et al. [2021], have analyzed social media text by combining Facebook conversations and survey data to identify substance use among homeless youth with high ROC-AUC scores.Segal et al. [2020], applied Word2Vec and Gradient Boosting Trees to medical insurance claims to analyze Opioid Use Disorder and achieved high sensitivity and specificity. Similarly, Ibrahimi et al. [2021], applied DNNs in identifying AUD cases from electronic health records, with much higher accuracy compared to the results from traditional algorithms. These studies underline substantial progress being made in predictive modeling for substance use disorders via diverse data sources and state-of-the-art techniques.

.

2.3 The Problem's Scope

Arguably, one of the most important and multi-dimensional current public health issues facing Bangladesh is the increasing tendency towards drug addiction among the ever-growing population of youths in this country. This paper will consider the socio-cultural, demographic, and health perspectives on the problem of drug addiction as it affects individuals and society. By focusing on peer influence, family dynamics, and conditions of mental health, the study tries to address the critical gaps in knowledge so as to provide actionable insights in combating addiction within the Bangladeshi context.

2.4 Challenges

This research faced difficulties at the data collection stage in terms of a complete dataset that represents most of the socio-demographic and behavioral patterns influencing drug addiction in Bangladesh. Extensive preprocessing addressed inconsistencies and partial incompleteness in the data. Other challenges included the generalization of models to changing demographics. The robust algorithms that minimize overfitting and improved the predictive accuracy of the solution include the Multilayer Perceptron, Extreme Learning Machine, and ensemble methods such as Blending and Voting. This made the models efficient for unseen data and hence applicable to real-world public health strategies.

CHAPTER 3

RESEARCH METHOLOGY

3.1 Proposed Methodology/Applied Mechanism

The provided diagram 3.1 outlines a structured methodology for detection of Drug Addiction using data processing and Machine Learning.

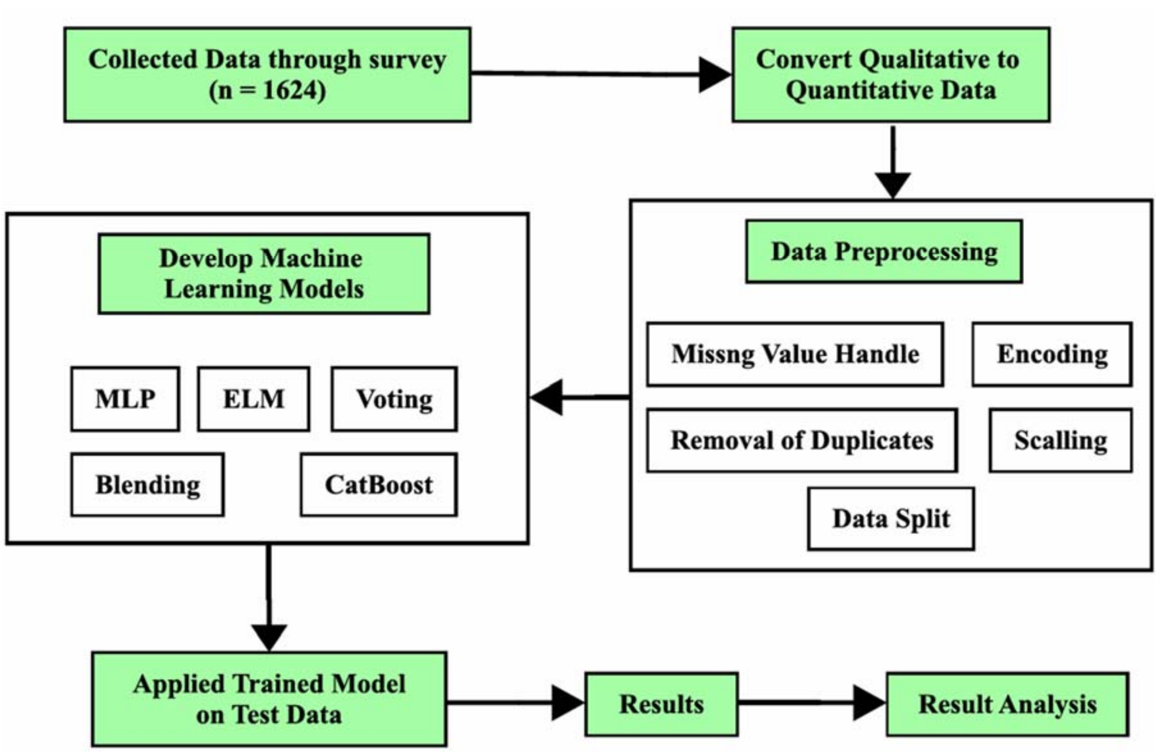

Fig 3.1: The working process to perform Drug Addiction Detection

Data for this study were collected through an online and face-to-face survey, targeting urban areas such as Dhaka and Sylhet, and targeting people aged between 15–27 years. In this survey, consideration was given to the socio-demographic and behavioral attributes of 1624 respondents. This dataset contains both addicted and non-addicted persons, so it is quite fair for analysis. Focusing on youth-a demographic most affected by addiction-means that this study is relevant to the public health challenges of Bangladesh.

3.2 Data Collection Procedure/Dataset Utilized

This study on drug addiction prediction has used the dataset collected through the survey targeted at youth-15-27 years-from Bangladesh. The survey encompasses socio-demographic, behavioral, and psychological attributes influential in determining the addiction tendency. In addition, it focuses on urban respondents from places like Dhaka and Sylhet to collect more data on increasing drug addiction among youth in these cities. This dataset represents a key primary source of data in finding the main patterns and predictors of addiction among highly vulnerable demographics.

The research methodology ensured a diverse representation of people from different educational and social backgrounds, with the majority of respondents being students. A structured questionnaire was designed to collect data on key variables like family dynamics, peer pressure, and mental health conditions, thus providing a comprehensive basis for analysis. Specific questions in the survey were aimed at eliciting information about addictive behaviors and their risk factors, hence allowing detailed insight into the socio-behavioral causes of addiction.

The dataset collected will represent an important source for training and validation of machine learning models developed in this work. Considering a representative youth demographic, with qualitative and quantitative variables, this dataset will provide new insights for the improvement of addiction tendencies and inform targeted public health interventions. The comprehensiveness of the data collected ensures a strong basis for predictive analytics and opens more avenues for future research into the domain of addiction prevention and rehabilitation.

After completing all the primary steps, the dataset's final uploaded images are shown in Figure 3.2.

	Timestamp	Age	Gender	Education (last/ongoing)	With whom do you live?	What is your thinking about drug?	With whom do you spend most of your times?	Did you ever fail in your life?	Do you get suicidal thoughts?	How is your family relationship?	...	Have you ever taken drug?	What is the name of the drug you used for the 1st time?	At what age you used drugs for the 1st time?	How often do you use drugs?	What drug do you use currently?	In what occasions do you use drugs?
0	8/15/2023 22:23	21.0	Male	Undergraduate	Family	Disease, should avoid	Family/Relatives	Yes	No	Very good	...	No	Never taken	No	No, I do not	No drug	Do not use
1	8/15/2023 22:27	22.0	Male	Undergraduate	Hall/Hostel	Should avoid	Alone	Yes	No	Satisfactory	...	No	Never taken	Never	No, I do not	No drug	Do not use
2	8/15/2023 22:29	21.0	Male	Undergraduate	Family	Disease, should avoid	Family/Relatives	Yes	No	Very good	...	No	Never taken	No	No, I do not	No drug	Do not use
3	8/15/2023 22:32	22.0	Male	Undergraduate	Family	Should avoid	Friends	Yes	Yes	Very good	...	No	Never taken	NaN	No, I do not	No drug	Do not use

Fig.3.2: Images of the uploaded data without encoding

3.3 Data Pre-processing

Extensive preprocessing was carried out to ensure the quality and suitability of the dataset for machine learning analysis. Missing data were treated by imputation techniques like statistical imputation and imputation using KNN in order to maintain data integrity. Synthetic data augmentation techniques were put to work for issues like class imbalance problems in addicted versus non-addicted cases. Normalization and scaling of numerical features were applied to unify the data, whereas encoding of categorical variables was done to seamlessly feed it into machine learning models.

After preprocessing, the data was divided into training and testing sets in an appropriate 70-20-10 proportion to guarantee robust model performance evaluation. Further EDA was performed in an attempt to visualize patterns, relationships, and outliers that may be contained in this dataset. EDA gave insight into the distribution of addicted and non-addicted subjects for some key features: age, mental health scores, and peer influence level. Such visualization provided critical insight that helped fine-tune model selection and parameter optimization. Qualitative data from the survey have been encoded into quantitative features by systematic techniques. Key predictors such as family dynamics, peer pressure, and attributes of mental health have been extracted and prioritized for machine learning modeling. This process of feature engineering ensured that the significant information relevant to the addiction risk analysis was kept while preparing the data for computational tasks. The Histogram of data frame of the pre-processed data are shown in Figure 3.3.

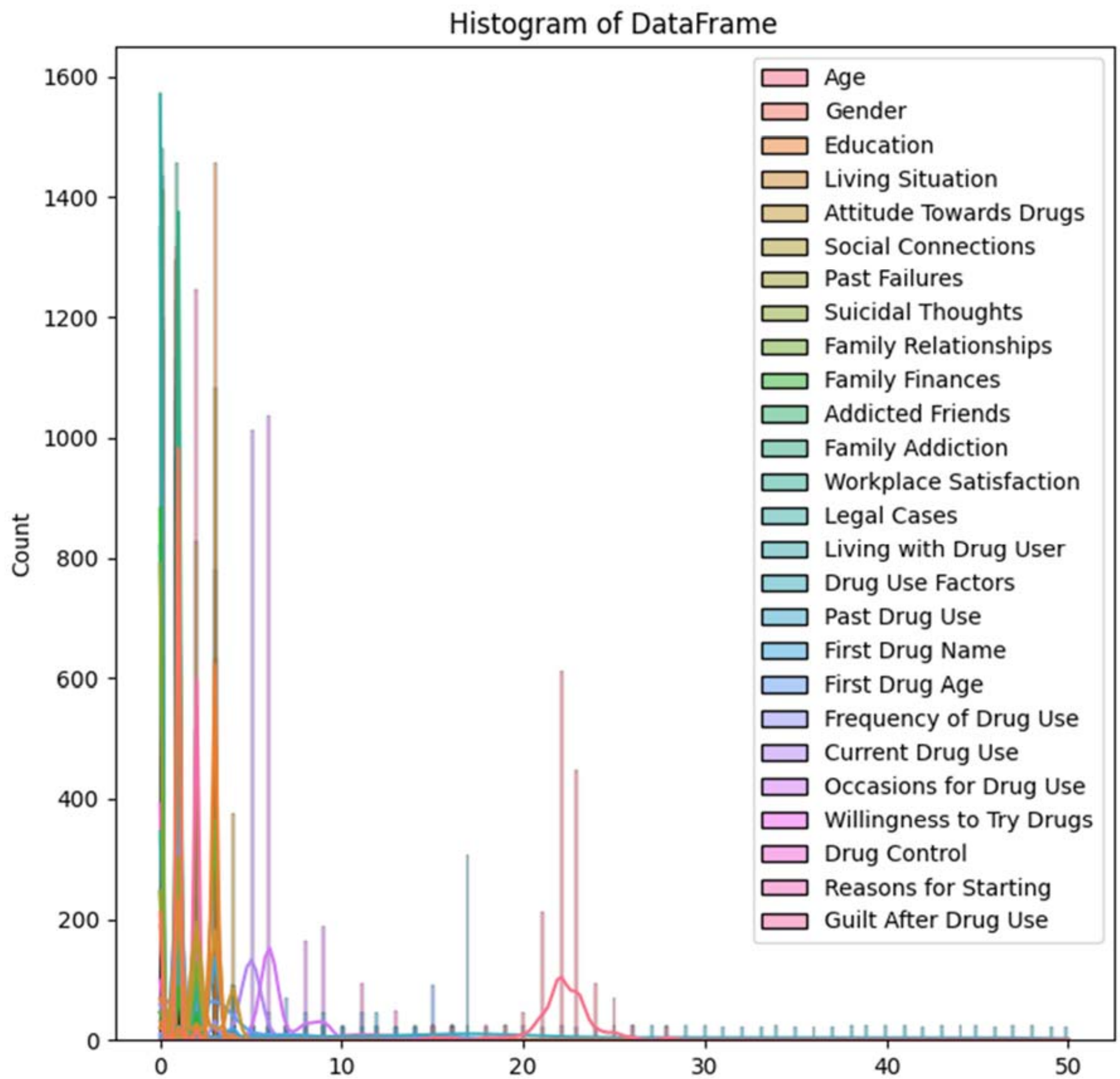

Fig.3.3: Histogram of DataFrame after completing preprocessing

3.3.1 Resizing Exploratory Data Analysis

Exploratory Data Analysis is done to get an overview of the dataset. EDA summarized the basic characteristics of data and made visualizations to show how the key attributes are distributed and varied. Some of the visualizations adopted in analysis include a bar plot for giving some statistical summary over the dataset. Above is the bar plot of mean, Std, 25%, 50%, 75%, and Max for different demographic and behavioral attributes. This plot helps in comparing the central tendency, dispersion, and range of the data, which are important to understand the distribution and variability within the dataset.

Analyses of these statistical measures outline patterns and anomalies that might lead to and confirm such concealed factors leading to addiction. For instance, the larger the standard deviation values, the greater variation in certain attributes, likely leading to higher risk factors. A similar look into the average and median could shed more light on data's central tendencies; thus, making it meaningful for us in understanding typical characteristic values of an addicted vs. non-addicted persons. The EDA process is an important preprocessing step in machine learning, enhancing model accuracy and reliability by the identification of key features and outliers to support public health policies and interventions. The Figure 3.3.1 shows the Bar plot diagram.

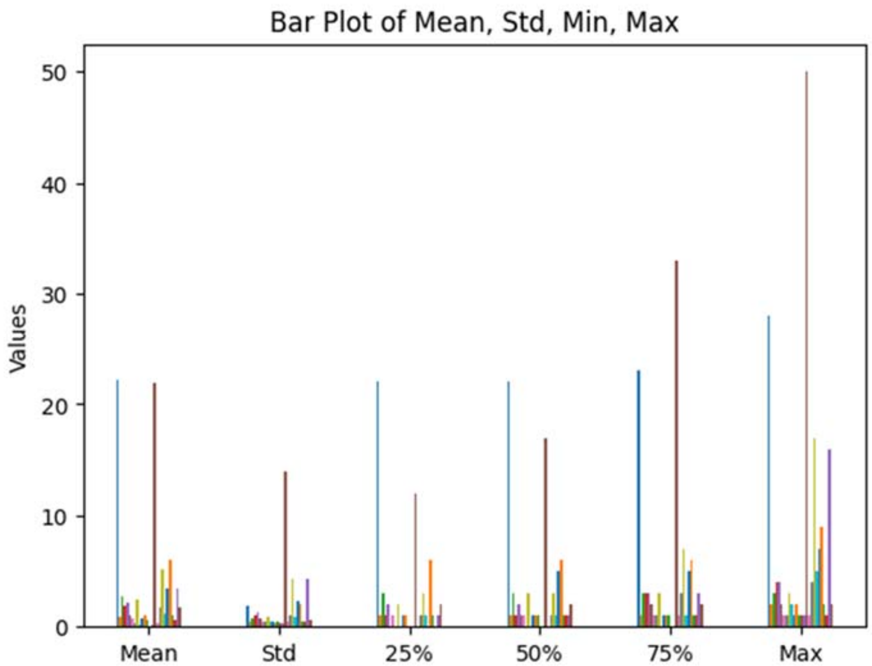

Fig.3.3.1: Bar plot diagram on Box Plot of DataFrame

3.4 Deep Learning Models

This research developed Multilayer Perceptron, Extreme Learning Machine, CatBoost, Blending (NN, GBM, KNN), and Voting (LR,RF,SVM) and then detected by the preset parameters.

3.4.1.Multiplayer Perceptron (MLP)

MLPs are feedforward artificial neural networks that consist of an input layer, one or more hidden layers, and an output layer. In the current work, MLP was utilized as a powerful classifier model in predicting drug addiction tendencies. Every layer in MLP is composed of fully connected nodes that adopt a weighted sum of their input with subsequent application of activation functions. This process has empowered the network to learn higher-dimensional and complex patterns of a dataset. Every node (neuron) in a layer has a weight that connects it to every other node in the layer below [18]. The MLP, being able to capture complex dependencies, gave an impressive result of 99.38% accuracy, one of the best in this work.

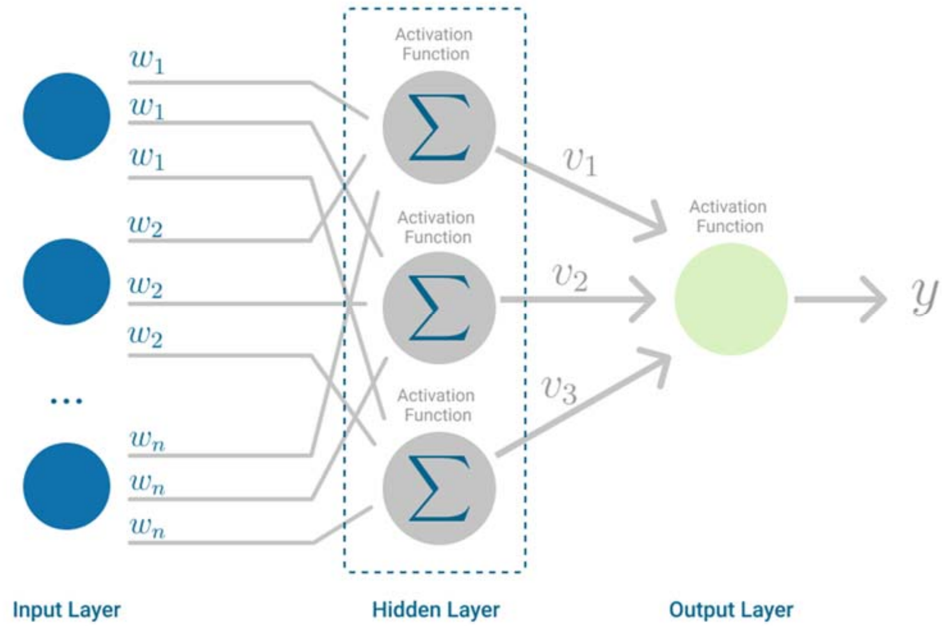

Fig. 3.4.1: Multilayer Perceptron Model Architecture (Google)

3.4.2Extreme Learning Machine (ELM)

The Extreme Learning Machine is a single-hidden-layer feedforward neural network designed for fast training and efficient learning. Different from the traditional neural networks, ELM initializes the weights of the hidden layer randomly and analytically determines the output weights with extremely low computational cost. ELM was applied in this paper to predict addiction tendencies using a balanced dataset. It has been designed by nature for speed in learning important features from socio-demographic factors to psychological attributes. This technique, though simple, scored a high 99.38% accuracy, matching that of MLP. This could be credited to the great efficiency in handling nonlinear high-dimensional data, making it one of the perfect candidates for any addiction prediction task. The output weights are determined analytically [19].

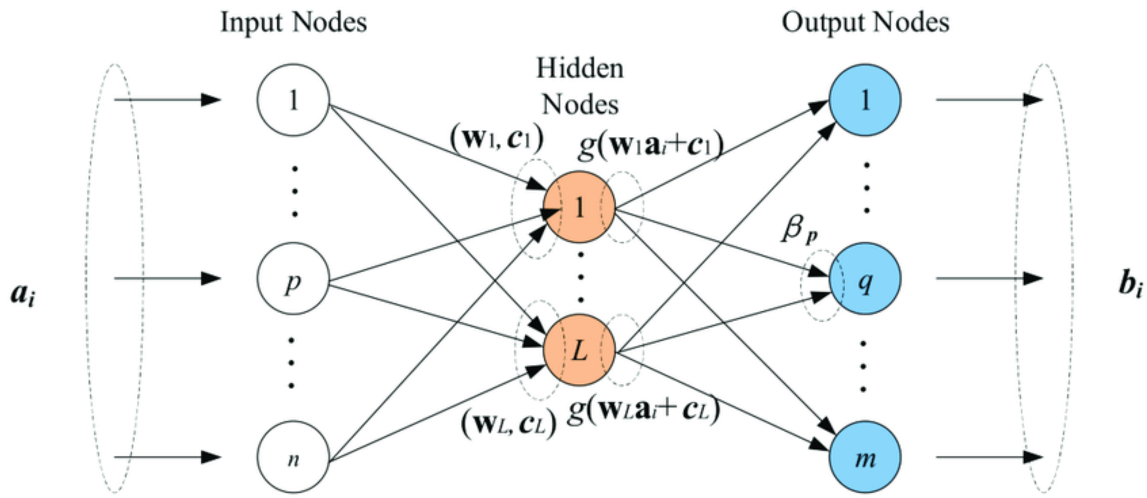

Fig.3.4.2: Extreme Learning Machine Architecture (Google)

3.4.3 CatBoost

In this paper, the CatBoost algorithm of gradient boosting was used because it is designed for improving predictive accuracy with reduced overfitting on categorical data. The algorithm applies ordered boosting and efficient handling of categorical features, hence suitable for analysis where there is a complex interaction between different variables. CatBoost then processed the socio-demographic and behavioral features for addiction prediction, learning hierarchical interactions and dependencies between them. The performance achieved a remarkable accuracy of 98.15% due to its capability in minimizing overfitting and efficiently handling categorical attributes of the data. This evidences its reliability in identifying addiction tendencies while keeping strong generalization. It is designed to improve the accuracy and speed of training by reducing overfitting and supporting ordered boosting [20].

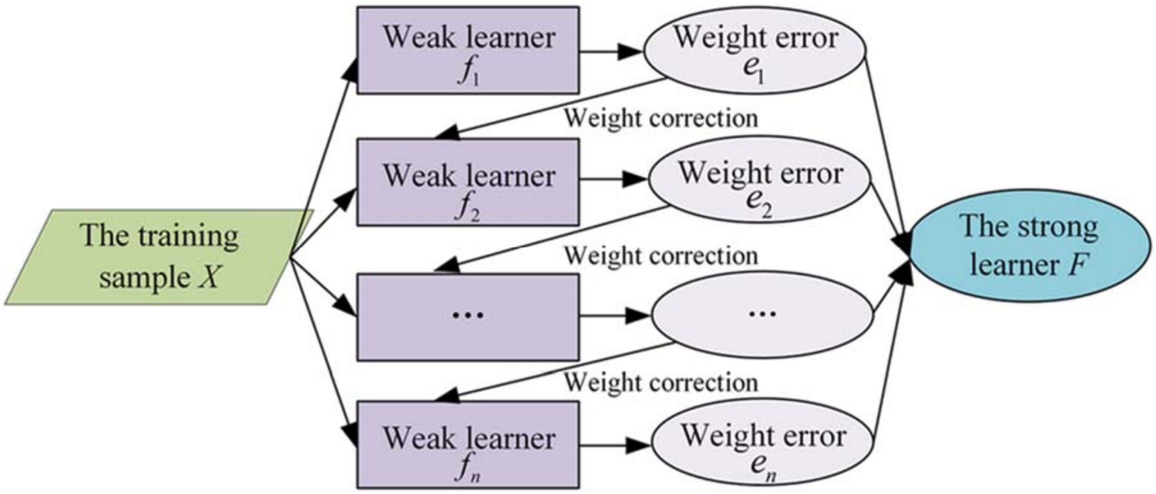

Fig.3.4.3: Extreme Learning Machine Architecture (Google)

3.4.4. Blending (Neural Networks + Gradient Boosting + KNN)

Blending: here, the ensemble model combines predictions coming from Neural Networks, Gradient Boosting, and K-Nearest Neighbor. It improves the classification performance. Each base model gets trained independently on the very same dataset, and their outputs are combined into a kind of meta-model-most of the time Logistic Regression, actually giving the final prediction. This ensemble approach allows the Blending model to take advantage of the strengths of different algorithms: the capturing of complex patterns with the Neural Networks, effective handling of nonlinear relationships with Gradient Boosting, and ensuring sensitivity to the local data structures with KNN. The combination of the predictions above allowed the Blending model to reach an accuracy of 98.15% with robust and reliable predictions. In this way, the Blending approach improved accuracy by integrating various algorithms, and general reliability of our classification system.

3.4.5. Voting (Logistic Regression + Random Forest + SVM)

A voting classifier was used in this study, combining the predictions using models by LR, RF, and SVM; each of these models was separately trained over the same dataset and was strong in its way. The final predictions, however, can be made through the Voting Classifier by either hard voting, or through soft voting. it computes the average probabilities to determine the final class. This thus effectively exploits the complementary strengths of individual constituent models in a more comprehensive manner that enhances both the accuracy and robustness of the results.

This will combine the predictions done by Logistic Regression, Random Forest, and Support Vector Machines to make a decision; hence, the Voting ensemble model. The ensemble shall use Hard Voting-majority rule, along with Soft Voting-averaging of predicted probabilities-to improve performance in class classification. This might be due to the strengths different algorithms bring into the Voting model: while Logistic Regression brings interpretability, Random Forest captures interactions between features, and Support Vector Machines do well in high-dimensional data spaces. Since it only yielded a moderately lower accuracy of 93.85% compared with the other models, a Voting ensemble remains comparably robust by integrating several diverse learning algorithms into one, thus presenting the base to an addictive prediction framework.

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Evolution Methods

Various performance matrices evaluate the machine learning classification models' blooming performance across a range of applications. The performance matrices listed below are taken into consideration:

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

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

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{iii}$$

$$\text{F1 Score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{iv}$$

4.2 Experimental Results & Analysis

The experimental results of this study provide evidence of the efficiency of the used machine learning models in the prediction of drug addiction tendencies. Among five models tested, Multilayer Perceptron (MLP) and Extreme Learning Machine (ELM) outperformed others with an impressive accuracy of 99.38%, precision of 100%, and recall of 98%, resulting in an F1 Score of 99%. While CatBoost and the Blending ensemble model were relatively close, at 98.15%, the robust nature of these models toward the intricate features of the predictors could not be concealed. Meanwhile, the Voting ensemble combines Logistic Regression, Random Forest, and SVM with the result of an accuracy at 93.85%, thereby giving a balance but still less accurate result. These results show the potential of advanced machine learning techniques, especially MLP and ELM, in classifying addiction tendencies based on socio-demographic and behavioral factors with a high degree of accuracy for targeted interventions and improvement in public health strategies. The purpose of the study is to check the performance of different machine learning algorithms in effectively predicting drug addiction. Experimental analysis involving five machine learning algorithms of evaluation on drug addiction tendency in Bangladesh using survey data was employed. Performance matrix of the models is given in Table 4.1.

Model	**Test Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score (%)**
MLP	99.38	100	98	99
ELM	99.38	100	98	99
CatBoost	98.15	94	100	97
Blending	98.15	94	100	97
Voting	93.85	88	91	90

Table 4.1: Finding the best result between the Result of Trained models

From this table, we can say that the MLP and ELM model is best suited to predict the drug addiction by using the preset parameter of these model. The performance evaluation of different machine learning models in predicting tendencies toward drug addiction has displayed the strengths and weaknesses of the models. Among the models analyzed, MLP and ELM showed the highest test accuracy of 99.38% each, indicating excellent classification performances for addiction tendencies. These models have also resulted in perfect precisions of 100%, high recalls of 98%, and an overall F1 Score of 99%. This is in indication of their strength in identifying both addicted and non-addicted subjects with least false positives or negatives.

Then CatBoost and the Blending ensemble model came next with 98.15% for the test accuracy. These models balanced precision and recall quite well, with both models scoring 94% for precision and 100% for recall, which further led to an F1 Score of 97%, hence their potential as useful tools for the analysis of addiction patterns, though at a level slightly lower than MLP and ELM.

The least accurate model, with 93.85% accuracy, is the Voting ensemble model that consists of Logistic Regression, Random Forest, and SVM. This gives a precision of 88%, a recall of 91%, and an F1 Score of 90%. Though it gives good performance, it is way behind other models in dealing with complex patterns in the dataset.

In summary, these results have shown the stronger predictive performances of MLP and ELM, as further supported by the good performances of CatBoost and Blending models. On the contrary, the Voting model has provided enough hints for further optimization in order to improve its applicability. These results really pinpoint the need to explore advanced machine learning techniques that will further improve understanding and prevention against drug addiction in Bangladesh. Here are some confusion matrix of all of the 5 developed models.

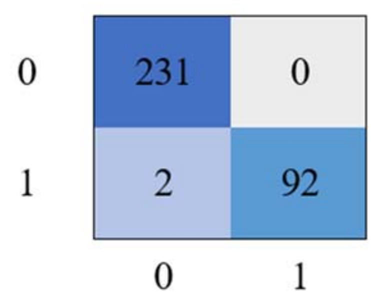

Fig 4.3.1. Confusion Matrix of the MLP model prediction

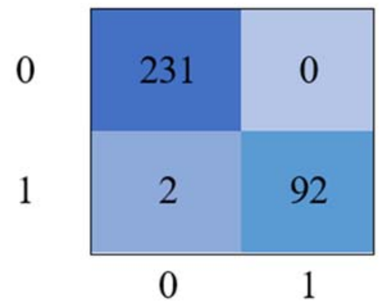

Fig 4.3.2 Confusion Matrix of the ELM model prediction

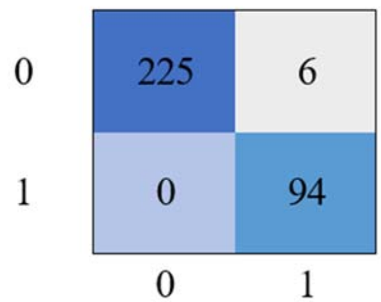

Fig 4.3.3. Confusion Matrix of the CatBoost model prediction

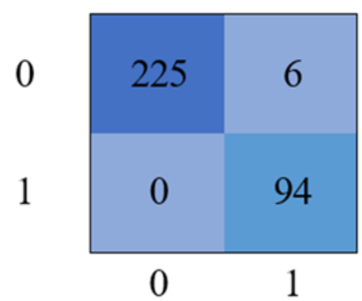

Fig 4.3.4. Confusion Matrix of the Blending model prediction

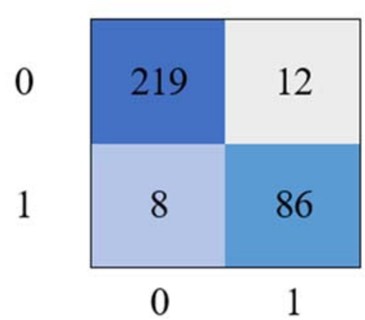

Fig 4.3.5. Confusion Matrix of the Voting model prediction

The confusion matrices of Figures 4.3.1, 4.3.2, 4.3.3, 4.3.4, and 4.3.5 show the performances of machine learning algorithms, namely MLP, ELM, CatBoost, Blending (NN+GBM+KNN), and Voting (LR+RF+SVM). One can easily observe that for well-characterized datasets with suitable model selection, the True Positives improved significantly. The size and attributes of the dataset also influenced the occurrence of False Positives, showing areas where the models excelled or had difficulties in classifying correctly instances related to addiction. This can be very useful in determining possible biases, tendencies to overfit, or further optimization areas. On the other hand, MLP and ELM models demonstrated better performance, which reveals their better capability of handling the complexity of the dataset.

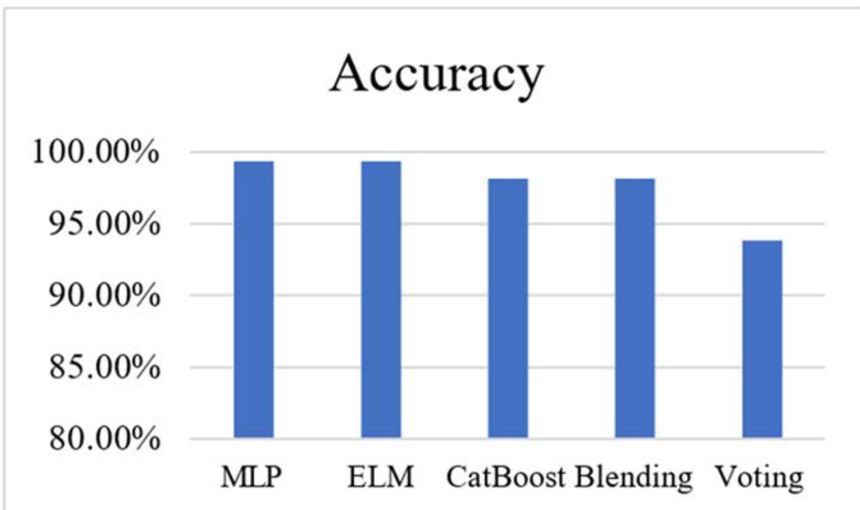

Fig 4.3.6. Comparison of the results between different models

Fig. 4.3.6 highlights the promising results of the predictive abilities of the selected machine learning models, notably MLP and ELM achieving 99.38% accuracy in identifying drug addiction tendencies from the survey data. Yet, limitations in data coverage and the complexity of the models emphasize the need for a cautious interpretation of these findings.

4.3 Discussion

In drug addiction prediction, the quality of a dataset and the performance of a predictive model are very key in informing the formulation of appropriate interventions. A strong predictive model requires a good-quality dataset as its backbone, from which meaningful patterns can be extracted to classify the addictiveness tendencies with precision. The following discusses how good-quality data and advanced models improve the predictive capability of drug addiction analysis.

First, a complete and balanced dataset allows the models to focus on the most relevant socio-demographic, behavioral, and psychological attributes. Key predictors in this study are family dynamics, peer pressure, and mental health conditions. With a representative sample in a diverse dataset, clear lines can be drawn between addicted and non-addicted persons by models such as MLP and ELM. This strong foundation reduces the biases and makes the model learn the patterns indicative of addiction tendencies rather than the noise.

High-quality data further improves the generalizability of machine learning models to real-world scenarios. For example, class imbalances between addicted and non-addicted individuals were tackled using synthetic data augmentation. While this increased the robustness of the models, it also spoke volumes about the integrity of data to be maintained so as not to overfit or misrepresent addictive behaviors. Models trained on reliable data can better predict addiction risks across diverse socio-economic and cultural contexts, thus aiding in the formulation of targeted interventions.

Besides, other sophisticated machine learning algorithms such as MLP and ELM have reported an accuracy of 99.38% in this work, proving their prowess in deciphering complex patterns within the data. These models leverage their architectural strengths in capturing intricate relationships between predictors, hence the precise classifications. However, their intricateness also presents a challenge in interpretability-a factor that needs to be addressed in translating the predictions into actionable public health strategies.

The ensemble models are Blending and Voting, which include the strengths of various algorithms, hence performing strongly. The Blending model, including Neural Networks, Gradient Boosting, and KNN, attained an accuracy of 98.15%. An ensemble model will ensure that different strengths of individual algorithms integrate for improved reliability and predictive power. These types of models are very important in enhancing the scalability and adaptability of addiction prediction systems across different regions.

More succinctly, high-value predictions to be realized in drug addiction demand high-quality data and state-of-the-art machine learning models. Both allow for the detection of critical addiction risk factors, improving model reliability and building region-specific public health strategies. As predictive analytics in addiction research continues to be refined, the relentless focus on data quality, model interpretability, and scalability will continue to serve as an instructive function in serving the greater good for society in this addiction crisis.

CHAPTER 5

IMPACT ON SOCIETY, HEALTH AND SUSTAINABILITY

5.1 Impact on Health

Machine learning for drug addiction among the youth of Bangladesh may be one of the most innovative applications with respect to public health. It can help in early identification of susceptible individuals, thereby reducing the health consequences associated with addiction, including mental illness, substance dependence, and chronic diseases. Predictive models further help in providing personalized intervention strategies, related to certain risk factors like family dynamics or peer pressure, which eventually improve both mental and physical well-being. The integration of such technologies within a public health framework can lower healthcare burden through shifts toward prevention and early care that minimize the need for extensive rehabilitation.

5.2 Limitations

Despite the valuable contribution of this study, a number of limitations are apparent. Firstly, most of the responses came from people in Dhaka and Sylhet, making it difficult to generalize these findings to other areas in Bangladesh with diverse socio-economic and cultural backgrounds. In addressing data imbalance, synthetic data augmentation was adopted, but a perfectly balanced distribution between addicted and non-addicted individuals is difficult to realize, which may affect the robustness of the model. This introduces possible biases into the dataset, as self-reported data cannot capture all the complexity and nuances of addictive behaviors. Besides, although the MLP and ELM models have attained a remarkable accuracy of 99.38%, their intricate architecture raises problems of interpretability, hence limiting immediate applicability for actionable policy insights. Potential work in the future might make efforts to overcome such limitations by using more diverse data, longitudinal studies, and cutting-edge explainable AI techniques for better usability and relevance of the findings.

5.3 Ethical Aspects

Machine learning in addiction research gives rise to a lot of ethical issues, particularly relating to data protection and the guarantee of equal opportunities. Trust and respect for ethical considerations among survey respondents require the guarantee of their anonymity and data security. Further efforts have to be taken for possible biases of the dataset in order to allow unbiased predictions in different demographic subgroups. Beyond the fact that the design of models should be open, interpretability gives trust to predictive outcomes and provides a responsible approach to the application of technologies.

5.4 Sustainability Plan

A sustainable framework of machine learning implementation in addiction prevention will involve continuous refinement of algorithms to adapt to emerging patterns and behaviors. Training programs for public health professionals and policymakers ensure that these tools are used ethically and effectively. By leveraging cloud-based technologies with energy-efficient practices, environmental impact is reduced while scalable applications are made possible. Such collaborations among governmental and community organizations help in setting up a strong infrastructural backbone for scaled-up adoption and integration within the national addiction prevention strategies through predictive analytics.

Long-term sustainability will require dynamic feedback in order to continuously monitor and evaluate the effectiveness of machine learning models in addiction prevention. This involves continuous gathering of data from changing behavior and updating the models to keep their accuracy and relevance high. Moreover, the collaboration with private technology companies could boost innovation in the use of resource-efficient AI deployment tools. Integrating these into education and workplaces can create even more awareness and facilitate early detection. Policies that focus on sustainable infrastructure and renewable energy sources will keep the ecological footprint of such systems at a minimum, reinforcing commitments to both social and ecological well-being.

CHAPTER 6

CONCLUSION ANDFUTURE WORK

6.1 Summary of the Study

This study adopted the application of machine learning in modeling and predicting drug addiction tendencies among Bangladeshi youth aged 15-27 years using survey data obtained from 1624 respondents. Some of the key findings obtained were that MLP and ELM yielded the highest accuracy of 99.38%, followed by CatBoost and Blending at 98.15%. The critical predictors identified were family dynamics, peer pressure, and mental health conditions, which highlighted the importance of these variables in assessing vulnerability to addiction. The ensemble techniques, Blending and Voting, showed that a combination of several algorithms indeed outperformed the rest in improving robustness and reliability within the predictions. This study highlights the potential of predictive analytics in driving early interventions and evidence-based public health strategies against drug addiction.

6.2 Conclusions

This research conducted an investigation into the various causes that lead the youth of Bangladesh to addiction to drugs by using the wide dataset obtained from survey data consisting of socio-demographic, behavioral, and psychological attributes of respondents. Advanced machine learning techniques such as MLP, ELM, and CatBoost have been used for the selection of important predictors of drug addiction related to family dynamics, peer influence, and other mental health disorders. Key findings were that MLP and ELM models showed excellent accuracy of 99.38%, proving to be very promising in the prediction of addiction tendencies. In turn, Blending and CatBoost are ensemble methods showing strong performance with an accuracy of 98.15%, indicating that machine learning has a promising future in addiction research. These results also indicate that predictive analytics is necessary for early intervention and tailored public health strategies in fighting drug addiction in Bangladesh.

6.3 Implication for Further Study

The findings of the present study give several ways in which future research may further understanding and the development of better prevention strategies on drug addiction in

Bangladesh. Further studies should investigate interactions between addiction and other behavioral and health problems, including the long-term effects on mental health, chronic diseases, and social functioning. Longitudinal studies are necessary in order to trace the progression of addiction, relapse, and the effectiveness of intervention across time. Advanced predictive models, such as deep learning and hybrid ensemble algorithms, may lead to better accuracy and extract more hidden patterns in the addiction risk factors. Further studies are also needed on geographical and socio-cultural variability in patterns of addiction to aid region-specific interventions and policies. These directions build upon the current findings and can inform more targeted and effective public health strategies.

The current study proposes a few major future research paths that would contribute to drug addiction prevention in Bangladesh. Further investigation is warranted to understand interaction between addiction and mental health or other chronic diseases and their long-term impacts on individuals' health and functioning in society. Longitudinal studies will be fundamental in tracking the course of addiction, relapse over time, and the impacts of interventions. Advanced models of prediction, including deep learning and hybrid ensembles, can further improve accuracy and find deeper patterns of addiction risk factors. Second, geographical and socio-cultural variations in addiction patterns would go a long way in advocating for region-specific interventions and policies. Increasing data to represent more diverse demographic groups and using multiple data sources may further enhance the generalizability of the models, thus yielding better public health strategies.

REFERENCES

[1] D. Kumari, S. Kilam, P. Nath. et al. “Prediction of alcohol abused individuals using artificial neural network”. Int. j. inf. tecnol. 10, 233–237 (2018).

[2] Singh, J., & Gupta, P. K. (2017). Drug addiction: current trends and management. Int J Indian Psychol, 5(1), 2348-5396.

[3] Ahad, M. A., Chowdhury, M., Kundu, I., Tanny, N., & Rahman. M. W. (2017). Causes of drug addiction among youth in Sylhet city of Bangladesh. IOSR Journal Of Humanities And Social Science (IOSR-JHSS), 22(5), 27-31.

[4] Beattie, M., & Nicholson, C. (2021). Feature extraction for heroin-use classification using imbalanced random forest methods. Substance Use & Misuse, 56(1), 123–130.

[5] Mary R. Lee, V. Sankar, A. Hammer, W. G. Kennedy, J.J. Barb, P. G. McQueen, L. Leggio, “Using Machine Learning to Classify Individuals With Alcohol Use Disorder Based on Treatment Seeking Status”, EClinicalMedicine, Volume 12,2019, Pages 70- 78,ISSN 2589-5370.

[6] Marcon, G., de Ávila Pereira, F., Zimerman, A., da Silva, B. C., von Diemen, L., Passos, I. C., et al., (2021). Patterns of high-risk drinking among medical students: A web-based survey with machine learning. Computers in Biology and Medicine, 136, Article 104747.

[7] Habib, S. E., Bhuiyan, A. Q., Hossain, H. T., & Lovejoy, F. H. (2001). Drug use and social circumstances: a study among different target groups in northwest Bangladesh. Nihon ArukoruYakubutsuIgakkaiZasshi= Japanese Journal of Alcohol Studies & Drug Dependence, 36(2), 107-123.

[8] Jing, Y., Hu, Z., Fan, P., Xue, Y., Wang, L., Tarter, R. E., et al., (2020). Analysis of sub- stance use and its outcomes by machine learning I. Childhood evaluation of liability to substance use disorder. Drug and Alcohol Dependence, 206, Article 107605.

[9] Y. Zhang, J. Liu, Z. Zhang, and J. Huang, "Prediction of Daily Smoking Behavior Based on Decision Tree Machine Learning Algorithm," 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC), Beijing, China, 2019, pp. 330-333.

[10] Segal, Z., Radinsky, K., Elad, G., Marom, G., Beladev, M., Lewis, M., et al., (2020). Development of a machine learning algorithm for early detection of opioid use disorder. Pharmacology Research & Perspectives, 8(6), e00669.

[11] Song, S., Zilverstand, A., Gui, W., Li, H. J., & Zhou, X. (2019). Effects of single session versus multi-session non-invasive brain stimulation on craving and consumption in individuals with drug addiction, eating disorders or obesity: A meta-analysis. Brain stimulation, 12(3), 606-618.

[12] Wadekar, A. S. (2020). Understanding Opioid Use Disorder (OUD) using tree-based classifiers. Drug and Alcohol Dependence, 208, Article 107839.

[13] C. Frank, A. Habach, R. Seetan, A. Wahbeh "Predicting Smoking Status Using Machine Learning Algorithms and Statistical Analysis", Advances in Science, Technology and Engineering Systems Journal, vol. 3, no. 2, pp. 184-189 (2018).

[14] Dou, Z. Y., Barman-Adhikari, A., Fang, F., & Yadav, A. (2021). Harnessing social media to identify homeless youth at-risk of substance use. Proceedings of the AAAI Conference on Artificial Intelligence, 35(17), 14748–14756.

[15] Sahu, K. K., & Sahu, S. (2012). Substance abuse causes and consequences. Bangabasi Academic Journal, 9(12), 52-59.

[16] S. Kinreich, J. L. Meyers, A. Maron-Katz. et al. "Predicting risk for Alcohol Use Disorder using longitudinal data with multimodal biomarkers and family history: a machine learning study". Mol Psychiatry (2019).

[17] Nasir, M., Summerfield, N. S., Oztekin, A., Knight, M., Ackerson, L. K., & Carreiro, S. (2021). Machine learning–based outcome prediction and novel hypothe- ses generation for substance use disorder treatment. Journal of the American Medical Informatics Association, 28(6), 1216–1224.

[18] Fong, D., Chu, T., Heflin, M., Gu, X., & Seneviratne, O. (2024). Predicting Depression and Anxiety: A Multi-Layer Perceptron for Analyzing the Mental Health Impact of COVID-19. arXiv preprint arXiv:2403.06033.

[19] Cheng, H. L., Tan, C. M., Chiu, C. M., Huang, H. Y., & Lee, Y. C. (2024). Using the ELM to Explore the Impact of Fake News on Panic Vaccination Intention: Taiwan's COVID-19 Vaccination Phenomenon. Journal of Global Information Management (JGIM), 32(1), 1-39.

[20] Fu, R., Yu, Z., Zhou, C., Zhang, J., Gao, F., Wang, D., ... & Yu, J. (2024). Artificial intelligence-based model for dose prediction of sertraline in adolescents: a real-world study. Expert Review of Clinical Pharmacology, 17(2), 177-187.

[21] Moonajilin MS, Kamal MKI, Mamun FA, Safiq MB, Hosen I, Manzar MD, Mamun MA. Substance use behavior and its lifestyle-related risk factors in Bangladeshi high school-going adolescents: An exploratory study. PLoS One. 2021 Jul 21;16(7):e0254926. doi: 10.1371/journal.pone.0254926. PMID: 34288956; PMCID: PMC8294555.

[22] Akter, M., Tofael Ahmed, M., Mazumder, S. H., Chowdhury, M. K., Shahriar Zaman Abid, M., & Rayan, M. A. (2024, May). Developing a Framework for Identifying the Scenario of Drug Addiction in Bangladesh. In International Conference on Information and Communication Technology for Intelligent Systems (pp. 243-251). Singapore: Springer Nature Singapore.

[23] Ahmed, F. U. (2024). Drug Control in Bangladesh: A Comparative Analysis Between Domestic Legal Mechanisms and International Standards (Doctoral dissertation, East West University).

[24] Waleed, M., Sajjad, M., Shazil, M. S., Tariq, M., & Alam, M. T. (2023). Machine learning-based spatial-temporal assessment and change transition analysis of wetlands: An application of Google Earth Engine in Sylhet, Bangladesh (1985–2022). Ecological Informatics, 75, 102075.

[25] Wang, H., Xiang, Z., Miller, D. J., &Kesidis, G. (2024, May). Mm-bd: Post-training detection of backdoor attacks with arbitrary backdoor pattern types using a maximum margin statistic. In 2024 IEEE Symposium on Security and Privacy (SP) (pp. 1994-2012). IEEE.

[26] Ahmed, F., Ahmed, F., Mubin, A. S., Rahman, H., & Ayon, R. (2024, June). Drug Addiction Analysis in Bangladesh Using Machine Learning. In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE.

[27] Sujana, S., Kumar, A. N., Charan, L. A. S., Nishitha, G., &Ratnababu, M. (2024, April). Comparative Analysis of Deep Learning Techniques for Drug Addiction. In 2024 IEEE 9th International Conference for Convergence in Technology (I2CT) (pp. 1-6). IEEE.

Drug_Addiction_Detection - Copy.docx

ORIGINALITY REPORT

12%	6%	4%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	4%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
3	Submitted to CSU Northridge Student Paper	1%
4	Dinesh Goyal, Bhanu Pratap, Sandeep Gupta, Saurabh Raj, Rekha Rani Agrawal, Indra Kishor. "Recent Advances in Sciences, Engineering, Information Technology & Management - Proceedings of the 6th International Conference “Convergence2024” Recent Advances in Sciences, Engineering, Information Technology & Management, April 24–25, 2024, Jaipur, India", CRC Press, 2025 Publication	1%
5	Sukhpreet Kaur, Sushil Kamboj, Manish Kumar, Arvind Dagur, Dhirendra Kumar Shukla. "Computational Methods in Science and Technology", CRC Press, 2024 Publication	<1%