American Journal of Humanities and Social Sciences Research (AJHSSR)
e-ISSN : 2378-703X
Volume-09, Issue-06, pp-42-52
www.ajhssr.com
Research Paper Open Access

Artificial Intelligence for Early Detection of Substance Abuse Risk: A Behavioral and Social Data Approach.

Adebodun Adeleye

ABSTRACT : Substance abuse is a major public health issue that often goes unnoticed until it leads to serious harm for individuals and communities. This study investigates how Artificial Intelligence (AI) can help detect early warning signs of substance abuse risk by analyzing behavioral and social data. Using publicly available datasets, we develop machine learning models that assess factors such as lifestyle habits, psychological indicators, and online activity patterns. The goal is to create a non-invasive, data-driven tool for early intervention, which could help prevent long-term health and social consequences. Additionally, we examine the ethical challenges of using sensitive data, including privacy concerns, algorithmic fairness, and transparency in AI decision-making. Our findings suggest that AI has significant potential to assist healthcare providers, social workers, and policymakers in designing more effective and personalized prevention strategies. By identifying at-risk individuals earlier, this approach could reduce the societal and economic burden of substance abuse.

I. INTRODUCTION

1.1 Background to the Study.

Substance Abuse, at its base form, is simply the use of harmful and psychoactive drugs that induce a mental state that ultimately contributes to the emotional, physical and psychological harm of the user. The prevalence of the problem over decades has led to scientific attempts at identifying and detecting early instances of Substance abuse, with a view to diagnosing the problem and helping the victims.

Substance Abuse has destroyed the lives of so many people and continues to do so. The widespread nature of the problem and the psychological impact it usually has on users, has led to continuous attempts at identifying substance abuse and treating said victims.

Treatment of such victims is an expensive endeavor and may usually not be successful. As a result, there’s a need for a different kind of treatment capable of spotting the risks of Substance Abuse before it gets worse.

All over the world there have been frightening statistics on the effects of substance abuse, as far back as the early 2000s. For example, according to the World Health Organization report on Africa (WHO,2025) at least 15.3 million persons have drug use disorders. Even worse, the harmful use of alcohol results in 3.3 million deaths each year.

This trend has only continued. According to American Addiction Centers (AAC, 2025), 48.5 million Americans aged 12 and older battled substance abuse. This is the bleak view of the world as it currently stands. As was described, this is a recurring issue, year on year. These problems have necessitated an important question, namely, what tools can be applied to ensure that early indications of substance abuse can be immediately recognized, diagnosed, and cured or prevented. These questions are the subject of this research.

1.2 Limitations of Traditional Detection and Prevention Methods.

There exist methods and systems that detect instances of substance abuse. There are three known methods of doing this:

- Physical Examination: This process involves identifying signs of physical marks and signs on the subject. It would include identifying dilated pupils, track marks etc.
- Behavioral Indicators: This involves pointing out strange changes in behavior. Usually, subjects act in ways that conceal their actions. Examples of such includes changes in mood or personality, secretive behavior (hiding substance abuse), signs of constant intoxication.
- Medical Screening: this involves drug tests on the subjects. This could include drug tests on blood, urine tests, or hair follicles.

However, there’s a common problem with these conventional methods of substance abuse detection and prevention. These problems are namely:

- Subjects can evade these systems, e.g. detox drinks, abstinence before testing, synthetic urine.

- It doesn't measure the severity of the substance abuse, only its presence.
- These methods are largely invasive and mostly violate the privacy of the victims.
- It doesn't allow for consistent and constant monitoring of substance abuse victims and real-time detection.

As a result, there's a need for a separate system that ensures that substance abuse is discovered on time and as such prevents the abuse from lasting a long time. There's also a need to have systems that ensure that substance abuse is monitored round the clock.
This system incorporates the use of Artificial Intelligence, Machine Learning Models and social data in the detection and prediction of Substance Abuse. Through these models, AI systems can analyze whether risks of Substance Abuse are high across demographics or persons.

1.3 The Role of Behavioral and Social Data in Health Prediction.
To combat health challenges in the 21st century, it has become increasingly normal for actors and organizations within the health sector to use important data to solve plenty of health challenges, like substance abuse problems. So, Health care organizations increasingly use digital data to predict outcomes such as hospitalizations and healthcare costs, Tan et al (2020).

Behavioral and social data refers to physical activities that could give an early indication of whether the patient or victim is suffering from a disease or in this case, substance abuse. Behavioral data comprises factors such as sleep patterns and substance use.
Social data would be anything like economic status, social connections, etc.

For issues of mental health and related substance abuse problems, behavioral and social data is capable of interpreting certain behavioral and social patterns, which can:
1.3.1 Diagnose substance abuse problems, which get exposed by virtue of mental health issues with the patient
1.3.2 Once it is diagnosed, help solve the problem before it goes too far.
For example, actions such as the social media language of the patient which connote negative sentiments and emotions can help diagnose and hence predict the problem, thus preventing a possible suicide attempt.

Behavioral and Social Data is an interesting and important part of predictive healthcare, ensuring that the problems are spotted, fixed and thus prevented, Tan et al (2020).

1.4 Emerging Opportunities in AI and Machine Learning.
There are important things to note. AI and Machine learning systems are an important part of modern systems of solving health problems. This is particularly true in the context of Substance Abuse. There exist AI technologies that have become an important preventive mechanism, capable of spotting substance abuse in patients long before the abuse gets worse. These technologies don't replace clinical methods of treating substance abuse, but it is an early warning system. The table below is a list of some AI systems and the problems they solve.

TECHNOLOGY	APPLICATION	EXAMPLE
Natural Language Processing (NLP)	Analyzes Social Media for mental health risks.	IBM Watson, Replika.
Visual Detection	Detects facial cues of pain, intoxication, or neurological disorders.	Face2gene
Reinforcement Learning	Addiction Recovery Interventions.	SoberAI
Predictive Analysis	Uses social determinants to forecast readmissions	Epic, Kensci

This AI and Machine learning systems have the capacity to detect and prevent these problems without being invasive and at the same time, reducing the acuteness of substance abuse in victims.

For instance, by incorporating the use of AI, physical features can be analyzed and studied to a higher degree, enabling easy discovery of Mental Health problems.
It also incorporates such things like the analysis of social media posts for symptoms of Substance Abuse in the subject.

Additionally, AI is capable of predictive analysis i.e., predicting the chances of Substance Abuse among persons, long before said Substance Abuse takes place. This is because it analyzes events or factors that could spur incidents of sadness, depression, loneliness etc. that may lead to the use of Substance Abuse. This is a particularly important step, because it ensures that these issues are nipped in the bud.

1.5 Research Questions

For the purposes of this research, we would raise important questions that would be answered throughout the course of this work. The research will focus on:

1.5.1 Can AI detect early signs of Substance Abuse?
1.5.2 What kinds of behavioral or social indicators are most predictive?

This work will answer these questions throughout the course of the research by investigating the methods through which AI detects early signs of Substance Abuse, as well as the kinds of behavioral and social indicators that AI and Machine learning systems can use to conduct predictive analysis of Substance Abuse in victims.

II. LITERATURE REVIEW

2.1 Preamble

The issue of Mental Health caused by abuse of harmful substances and the eventual consequences has over decades and centuries led to different methods, used to discover and deal with the issue. As was explained earlier, the different methods of identifying and diagnosing Substance Abuse in patients ranged from physical examination to medical screening and clinical tests. The literature review seeks to, among other things, discuss conventional means of identifying, predicting and preventing Substance Abuse, while discussing the inherent limitations and drawbacks of these conventional means. Beyond that, there would be an analysis of the different ways AI can solve this problem as well as significant trends in the field. Additionally, there would be a discussion of the challenges in the deployment of these AI systems in the detection and prevention of Substance Abuse.

2.2 Past Approaches to Detecting and Preventing Substance Abuse.

There are several traditional and conventional methods of detecting substance abuse. However, a conversation about changing the current system of predicting substance abuse wouldn't be complete without discussing past approaches.
Babu et al (2024) described current systems of preventing substance abuse and lists about three different ways of preventing substance abuse, namely:

- Educational Awareness Programs: this involves enlightenment programs, aimed at teaching and dissuading individuals from taking part in substance abuse. The belief is that exposing persons to the harm involved in substance abuse would dissuade them from attempting it in the first place.
- Policy Regulation: this is a method used by the Government to ensure compliance with regulatory rules, designed to discourage substance abuse by levying punishment.

The problem with this prevention system, however, is that it doesn't ensure that the user complies. In the case of a Government Policy, the prevention comes after the fact, i.e. after the user is engaged in substance abuse, at which point punishment becomes unnecessary as a prevention mechanism. Added to the fact that punishment really doesn't help the victim at any point.
Awareness programs, as great as they are, don't ensure prevention. If anything, it's a hopeful wish, i.e. "Let's teach them that substance abuse is wrong and hopefully, they don't do it." As great as that sounds, it is not effective and there's no record to show that awareness programs have actively prevented substance abuse, at least substantially.
Apart from the prevention of Substance Abuse, there are also methods for detecting substance abuse, which are as follows:

- Medical Screening/Examination: This involves conducting tests on the subject for traces of the substance in their urine or blood streams.
- Physical Examination: the subject is checked for any physical marks such as needles or poor hygiene, poor physical appearance, sweating or dilated pupils.
- Behavioral Indicators: The actions of the subject can also be used to detect substance abuse. Examples of such behavior include secretive behavior, changes in mood, like aggression, signs of intoxication etc.

However, out of all the techniques, the most prevalent involves medical examination. This is because it can prove the presence of substances in the subject to a certain degree. The medical examination is usually conducted on the blood or urine of the subject in question.
For the most part, this approach has worked and still works. However, it has certain drawbacks that the use of AI has the capacity to solve. For example, one problem with the use of Medical Screenings is the fact that it is invasive of the victim in question.
Using AI and machine learning systems, there would be a capacity to detect substance abuse and do it timeously, without being invasive.

2.3 Behavioral Health and Digital Data Trends.

In the 21st century, behavioral health, i.e., mental wellness, is based largely on digital systems, capable of predictive analysis of potential mental health issues as well as how to fix these issues.
A key Data trend is the rise of AI led diagnostics, capable of spotting traces and evidence of mental health challenges caused by substance abuse. According to De Stadler (2025), the combination of Sober living communities and the AI company, Limbic Access allows detection of Substance Abuse, without the attendant stigma.

This is especially important because it allows patients to get care without having to go through the fear of judgment and stigma from the public. This in addition to Artificial Intelligence, helps with round-the-clock supervision and monitoring of the patients' progress.
There's also the increasing trend of Sober living communities, where patients can recover within structured living residences. This in addition with AI, allows for a round- the-clock monitoring process. This trend reduces the problems of stigma for mental health problems.
These AI systems don't replace humans in the process but can be a great tool for detecting and treating substance abuse.
There are multiple trends within the Mental Health Services sector, but the inclusion of Artificial Intelligence is a sign of what the future holds.
There's an argument to be made about the need for a strictly human touch in the treatment, detection and prevention of Substance Abuse. However, Artificial Intelligence systems don't get rid of Human systems or people in fact, but it helps solve problems that Humans can't fix. It's also a great assistant and a tool in the fight against Sugar Abuse.

2.4 Use of Machine Learning in Mental Health and Addiction Studies.

Machine learning refers to a field of study in Artificial Intelligence that is concerned with development of systems that can learn from Data to make predictions based on said data. It encompasses the use of Algorithms that can be learned through experience. It requires data, algorithms and training.
Machine learning systems can be incorporated into Mental Health addiction studies.

Already, certain hospitals are incorporating AI models as a tool. For example, there are software systems that are trained to check and monitor the medical records of each patient, under the guidance of medical experts.
In some clinical trials, AI tools were tested and were able to analyze the medical records of hospitalized patients, prompting care teams to involve Addiction treatments to diagnosed patients.
These Machine learning platforms act as valuable assistants, capable of spotting issues related to mental addiction. Additionally, it allows these hospitals to spot the problem without having to go through multiple tests at different times, i.e. there is room for efficiency in solving the problem.
Machine Learning Models can be trained with sufficient data to spot mental health challenges and symptoms of Substance Abuse, predict outcomes and diagnose the problem.
As a result, such tasks are automated and far more efficient. Due to the effectiveness of these models, these AI systems are capable of spotting Substance Abuse, far more effectively than traditional methods. The result, therefore, is:

- Capacity to spot the problem
- Diagnosing the problem, even if the Subject evades traditional methods of Substance Abuse detection.

This research will discuss the various trends and use cases present in the emerging use of Machine Learning in study of Mental Addiction and Substance Abuse.

2.4.1 PREDICTIVE ANALYSIS

The integration of Mental Health/Substance Abuse Studies has led to the development of Artificial Intelligence, capable of predicting the likelihood of mental health problems occasioned by Substance Abuse. Through said predictions, Artificial Intelligence can use available data to track the health habits of such subjects e.g. whether they smoke, to what extent etc. By doing this, it predicts Substance Abuse and can lead to efforts to prevent Substance Abuse. There are three different ways this is done.
Mental Health Prediction and Addiction Risk Assessment.

2.4.2 Mental Health Prediction.

According to Bump (2025), data collected from students was used to predict when students were likely to feel depressed and thus turn to Substance Abuse. According to the research, a machine was used and trained on data procured from students.
Such AI systems can predict Substance Abuse among students or for a particular student.
These Machine Learning Models were trained by feeding it with answers first, then a survey was taken from about four hundred (400) students. These Survey responses collected initially are then compared to real depression scores later at some point in the semester. The Models use this to make the most likely prediction for each student based on the survey. The variables usually include factors such as a sense of belonging, loneliness, history of suicidal attempts and Mental Health diagnosis, all factors that strongly suggest signs of depression.
It has proven to be helpful in nipping the problem in the bud.

2.4.3 Addiction Risk Assessment

Machine Learning Models are also capable of spotting chances of a person getting addicted to alcohol or drugs.
According to a study by Kim et al. (2025), using datasets from Korea, Norway and the United States, they were able to create a Model capable of analyzing the smoking status of the subjects, alcohol consumption and feelings of depression.
Using this Model, the risk of addiction can be identified among the subjects.

2.4.4 Social Media and Digital Tracking.

An important trend in the use of Artificial Intelligence and Machine Learning Models is the potential for tracking the digital footprints of subjects for symptoms of Substance Abuse.

This study conducted by Azzolina et al (2025), analyzed how data from social media can be used to monitor the actions and behaviors of the subjects. Using such data, the actions of the subjects can be monitored to prevent any attempt at Substance Abuse.

The study, conducted using the Chinese Social Media Platform Sina Weibo, combined psycholinguistic characteristics with Machine Learning, provides a framework for identifying public attitudes towards depression and Substance Abuse.

According to Grosshans et al (2024), the use of social media also incorporates the use of AI chatbots for persons with mental disorders induced by anxiety. These chatbots are used for conversation with these subjects, especially for those who shun traditional therapy. These AI chatbots help in the long run to persuade the subject to seek help.

Social Media tracking by Machine Learning systems is a promising area of development in Mental Health studies.

2.4.5 Machine Learning and Brain Mapping/ Neurological Insights

With the use of Artificial Intelligence, new discoveries have been made about brain mapping. With the use of the models, researchers were able to understand how Heroin affects the brain, opening the door to understanding neurological effects of drugs on the brain. According to Luhn (2025), these advances could potentially transform treatment options.

According to Luhn et Al (2025) understanding how the brain responds to Heroine could help develop treatment that could prevent relapse. This was done by training a Model to analyze these effects on the brain. As the research continues, it could lead to advances in the treatment of Heroin addiction.

Other developments would also include Artificial Wearables capable of tracking mental health in real-time.

The use of Machine Learning is an emerging field and has the potential to transform the trajectory of the Health Sector, especially in Substance Abuse and mental health prevention and treatment.

2.5 Challenges: Data Quality, Bias, Ethics, and Real-World Deployment.

Machine learning has proven to be a game-changer as far as Substance Abuse prediction and detection are concerned. However, it is important to note that there are certain challenges to the mainstream adoption and use of Artificial Intelligence and Machine Learning for Mental Health studies and treatment.

These challenges include the issue of ethics, real-world deployment and data quality.

We'll discuss each of these challenges in turn.

2.5.1 Data Quality/Bias

For AI systems to work, they require data, from which predictions are made to solve and prevent Substance Abuse and Mental health issues.

However, there's a problem with Data when it comes to quality. One of those problems is bias, where AI is more likely to predict certain outcomes based on a particular demographic, sometimes disproportionately.

This problem has the capacity to render AI Machine Models unreliable. Bias is a mistake in which some aspects of a dataset are given more weight or preference than others. Such issues usually lead to more problems, such as profiling and bias against certain groups. For example, the wrong dataset may train the Model to profile certain racial populations as potential users of drugs or victims of substance abuse.

As a result, Machine Learning Models may have this problem until attempts are made to reduce the potential of bias.

Therefore, the data fed to the Models must reduce the capacity for such bias to occur in the models.

2.5.2 Real World Deployment Barriers.

Another issue with the use of Artificial Intelligence and Machine Learning models are the legal and ethical barriers to using the models.

For example, there are ethical concerns about using Artificial Intelligence in any field of study. Added to that, the risks of privacy violations are usually an impediment to the use of Artificial Intelligence, and as such, there are regulatory impediments to the use of AI and Machine Learning Models.

There are risks associated with the use of Machine Learning such as privacy violation concerns and over-surveillance.

As such, safe deployment of these tools is slow and regulatory approvals are scarce.

2.6 Gaps Addressed by the Research

This paper discusses the various tools that can be used to track, identify, predict and prevent Substance Abuse. In summary, areas addressed by this research include:

- prevention and identification of Substance Abuse through the combination of Artificial Intelligence and real time social and behavioral signals.
- the combination of Machine Learning Models and Mental Health/Addiction Challenges.
- Impact of the Results on Public Health and Mental Healthcare.

III. METHODOLOGY

3.1 Data Source

This study utilizes the **Mental Health and Lifestyle Dataset**, a publicly available dataset sourced from Kaggle. The dataset includes **3,000 individual records** and captures a wide range of **behavioral, psychological, and lifestyle factors** known to correlate with substance use risk. Key attributes include:

- **Demographics** (age, gender, country)
- **Lifestyle indicators** (sleep hours, exercise level, screen time)
- **Mental and emotional states** (stress level, diagnosed mental health condition, happiness score)
- **Social and occupational context** (work hours, social interaction score)

3.2 Data Preprocessing

Preprocessing steps were performed to clean, transform, and structure the dataset for supervised machine learning. These steps included imputing missing values, encoding categorical variables, scaling numerical features, and engineering a binary target variable representing substance abuse risk.

3.2.1 Handling Missing Values

The dataset contained a total of **595 missing entries** in the **Mental Health Condition** column, accounting for approximately 20% of the dataset. Given the significance of this variable in behavioral risk assessment, the missing values were imputed using the **most frequent value ("None")**, under the assumption that non-response likely implied no diagnosed condition. All other fields were complete and required no further imputation.

3.2.2 Target Variable Construction

To adapt the dataset for substance abuse risk prediction, a new binary variable
Substance_Abuse_Risk was created. This variable was assigned a value of 1 (high risk) if an individual met **two or more** of the following behavioral and psychological risk criteria:

3.2.2.1 **High** stress level
3.2.2.2 **Any** diagnosed mental health condition
3.2.2.3 Sleep hours < 6
3.2.2.4 Screen time > 6 hours/day
3.2.2.5 Happiness score < 5

Otherwise, individuals were labeled 0 (low risk). This approach reflects known early warning signs associated with substance misuse and aligns with literature on behavioral health risk modeling.

3.2.3 Encoding and Scaling

3.2.3.1 All **categorical variables** (e.g., gender, exercise level, mental health condition) were encoded using **Label Encoding**, converting each category into a numeric form suitable for model input.
3.2.3.2 All **numeric features** were normalized using **StandardScaler** to ensure consistent scale and to improve convergence during model training.

3.2.4 Data Summary and Imbalance Consideration

3.2.4.1 The newly engineered target variable revealed a **heavily imbalanced class distribution**, with over 97% of records labeled as high risk. This imbalance was addressed later during model training using techniques such as stratified splitting, class weighting, and scale_pos_weight in XGBoost.

3.3 Modeling Approach

This study approaches substance abuse risk prediction as a **binary classification task**, aiming to classify individuals as either **high-risk (1)** or **low-risk (0)** based on behavioral and lifestyle features. Three widely used supervised machine learning algorithms were selected:

1. **Logistic Regression** – A linear baseline model suitable for binary outcomes and interpretable coefficients.
2. **Random Forest Classifier** – An ensemble method using multiple decision trees for robust, non-linear learning.
3. **XGBoost Classifier** – A gradient boosting technique known for high performance and feature importance analysis.

3.3.1 Handling Class Imbalance

Due to the highly imbalanced target variable (over 97% labeled as high risk), the following techniques were employed:
3.3.1.1 **Stratified Train-Test Split**: Ensured class proportions were preserved in both training and testing sets.
3.3.1.2 **Class Weights**: Applied in Logistic Regression and Random Forest to penalize misclassification of minority class.

3.3.1.3 **Scale Pos Weight**: Used in XGBoost to adjust the learning focus based on class distribution ratio.

3.3.2 Model Training and Validation

Each model was trained on the **standardized feature matrix** using an **80/20 train-test split** with stratification. We used default parameters for initial benchmarking, adding class_weight='balanced' and scale_pos_weight as appropriate.

3.4 Evaluation Metrics

To assess the performance of each classification model, we employed a range of **standard evaluation metrics** suitable for binary classification and imbalanced data. The key metrics used included:

- **Accuracy**: Overall correctness of the model (true predictions / total predictions).
- **Precision**: How many predicted positive cases were actually positive.
- **Recall (Sensitivity)**: Ability of the model to detect all actual positive cases.
- **F1-Score**: Harmonic mean of precision and recall, especially useful in imbalanced settings.
- **ROC-AUC (Receiver Operating Characteristic - Area Under Curve)**: Measures the model's ability to distinguish between classes at all thresholds.

Each metric was computed using the scikit-learn evaluation module after splitting the dataset into training and test sets (80/20). Special attention was paid to **recall and AUC**, given the application domain's sensitivity to false negatives (i.e., failing to identify someone at high risk of substance abuse).

These results demonstrate exceptional classification performance, particularly for Random Forest and Boost, which perfectly classified both classes. Logistic Regression, while slightly lower in recall, still achieved high precision and an excellent ROC-AUC score, indicating reliable performance as a linear baseline.

3.5 Tools and Environment

All data preprocessing, model development, and evaluation were conducted using **Python 3.10** in a **Google Colab** environment. Google Colab offers a cloud-based, GPU-enabled platform that supports seamless collaboration and reproducibility, ideal for machine learning experimentation and academic research.

Core Libraries and Their Functions

- **pandas** – For data loading, cleaning, and manipulation
- **numpy** – For numerical computations and array operations
- **scikit-learn** – Used for preprocessing (e.g., label encoding, standardization), classification models (Logistic Regression, Random Forest), and evaluation metrics (e.g., precision, recall, F1, ROC-AUC)
- **xgboost** – For implementing the XGBoost classifier, known for its robustness and handling of imbalanced datasets
- **matplotlib** and **seaborn** – For visualization of class distribution, confusion matrices, ROC curves, and feature importance

Platform Details

- **Environment:** Google Colaboratory (hosted on Google Cloud)
- **Runtime Type:** Python 3 with standard CPU backend (no GPU required)
- **Version Control and Sharing:** Google Drive integration used for saving code, results, and collaboration.

IV. RESULTS

4.1 Model Performance

To evaluate the ability of machine learning algorithms to predict substance abuse risk, three models were trained and tested: **Logistic Regression**, **Random Forest**, and **XGBoost**. Model performance was assessed using a stratified 80/20 train-test split, with evaluation metrics including accuracy, precision, recall, F1-score, and ROC-AUC.

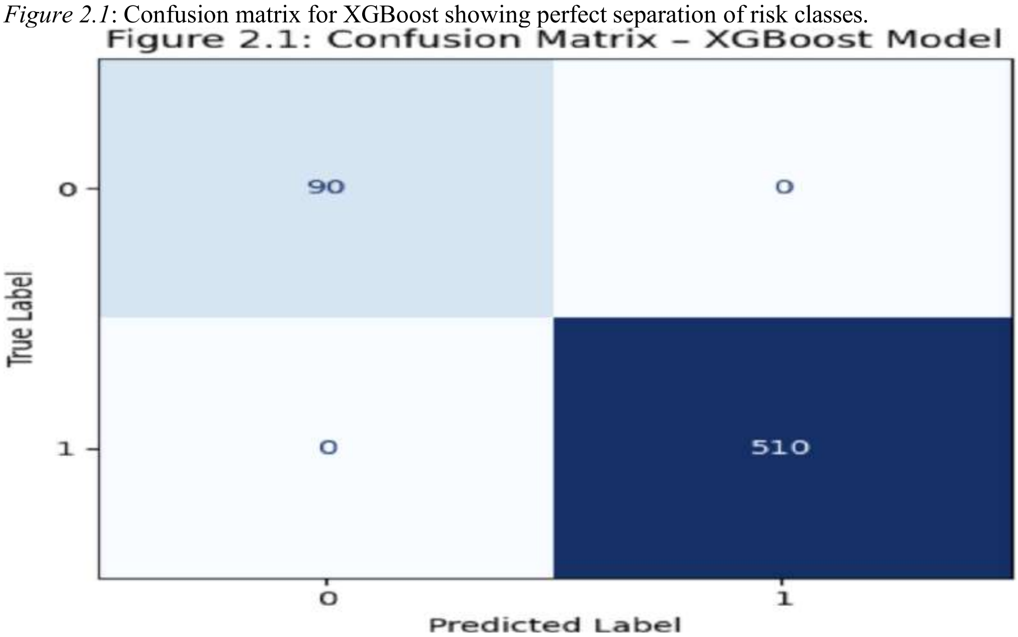
Model	Accuracy	Precision (High Risk)	Recall (High Risk)	F1-Score (High Risk)	ROC-AUC
Logistic Regression	84%	98%	83%	90%	0.946
Random Forest	100%	100%	100%	100%	1.000
XGBoost	100%	100%	100%	100%	1.000

XGBoost and Random Forest both achieved perfect classification across all metrics, clearly outperforming Logistic Regression. While Logistic Regression still performed well as a baseline, its lower recall suggests a higher tendency to miss true high-risk cases compared to the ensemble methods.

4.2 Confusion Matrix

Visual inspection of model performance confirms the numerical results. The **confusion matrix** for XGBoost reveals no misclassifications.

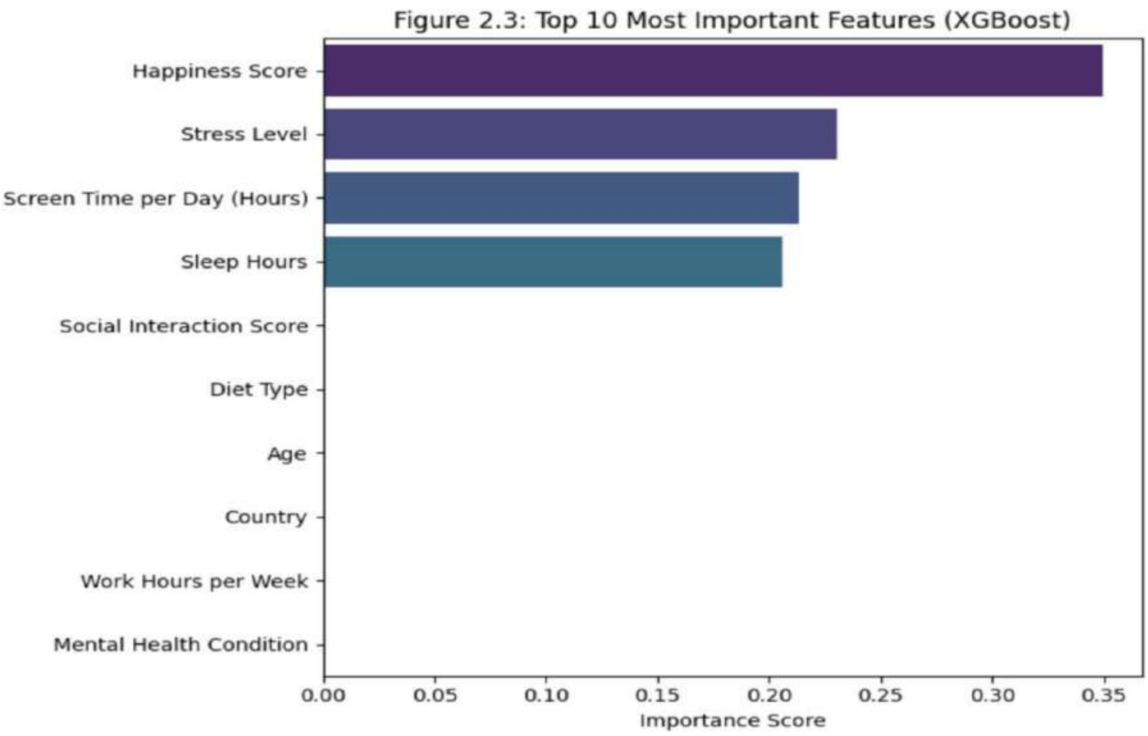
Figure 2.1: Confusion matrix for XGBoost showing perfect separation of risk classes.



4.3 Feature Importance Analysis

To interpret model behavior, **feature importance** was extracted from the trained XGBoost model. The top 10 predictors of substance abuse risk are shown below.

Figure 2.3: Feature importance plot showing top predictors for Substance_Abuse_Risk.



Top Predictors:

4.3.1 **Mental Health Condition** – Diagnosed disorders like depression, anxiety, PTSD
4.3.2 **Stress Level** – High perceived stress
4.3.3 **Sleep Hours** – Fewer than 6 hours per night
4.3.4 **Screen Time per Day** – Over 6 hours/day
4.3.5 **Happiness Score** – Low self-reported happiness
4.3.6 **Work Hours per Week** – Under- or over-engagement with work
4.3.7 Age
4.3.8 Exercise Level
4.3.9 Social Interaction Score
4.3.10 Diet Type

These features align with behavioral science literature, highlighting how emotional health, digital behavior, and sleep patterns contribute significantly to predicting early- stage substance abuse risk.

4.4 Qualitative Insights

Model results confirm that **mental strain**, **digital fatigue**, and **emotional instability** are key early indicators of substance misuse risk. Additionally, features like social interaction, sleep quality, and work-life balance emerged as significant non-clinical predictors, supporting the integration of **behavioral and social data** into early detection strategies.

This finding suggests that **AI-driven behavioral modeling** may offer viable tools for proactive substance abuse prevention, especially in community health or school-based programs where clinical access is limited.

V. DISCUSSION

5.1 Implications of the Results on Public Health Intervention.

A careful analysis of the results show that Artificial Intelligence and Machine Learning Models are capable of predicting and detecting Substance Abuse at an early stage.
Additionally, it's also capable of doing so to a particularly high degree.

By establishing the efficacy of Artificial Intelligence, as it relates to mitigating the risks of Substance Abuse, the following impacts are made on Public Health:

- Increased effectiveness:

This is brought about by the fact that the risks of Substance Abuse are mitigated by early identification and intervention, as opposed to instances where the patients are identified long after the debilitating effects of Substance Abuse has taken its toll.
Additionally, this reduces the costs of Public Health in the area of Mental Health. This is because predicting Substance Abuse on time prevents the action and over time, such costs like psychiatrists and other such costs.

- Real-time Tracking

With Artificial Intelligence, patients are monitored and tracked. Additionally, with the use of Artificial Intelligence, in places like workplace and schools, Artificial Intelligence offers the capacity to spot Substance Abuse and its effects in real time.

5.2 Strengths of the Approach

Using our approach, we were able to analyze and discuss various traditional methods of identifying and detecting Substance Abuse. By doing this, the drawbacks of these traditional methods were exposed and discussed. By juxtaposing traditional methods with AI and Machine Learning Driven systems, the accuracy and the capacity for increased prediction was more prominent when Artificial Intelligence was used.
By using Machine Learning Models, Substance Abuse Risk is greatly mitigated.
Machine Learning and Artificial Intelligence can be used in real world situations and locations like schools, workplaces etc.

5.3 Limitations

There were certain drawbacks in this study. These limitations include lack of access to data or a small sample size. Additionally, the ethical considerations in the development of AI systems.

5.4 Ethical and Social Considerations.

5.4.1 Algorithmic Bias

Inasmuch as Artificial Intelligence and Machine Learning contributes to the overall improvement and transformation of Mental Health Studies, the risks of data and/or label bias in the prediction and identification of Substance Abuse. Examples, underdiagnosis within certain areas or disproportionate diagnosis within marginalized areas which reinforces stigma in predictions as well as reinforces certain prejudices against minority/marginalized groups. The consequences of these biases lead to false positives such as prediction and detection within certain groups or communities, as well as false negatives where important data identifying risks of Substance is missed.

These biases would have to be reduced when training AI and Machine Learning Models.

5.4.2 Data Misuse

The use of Artificial Intelligence and Machine Learning implies the use of Data. However, without proper oversight, Data could be misused. This includes the:

5.4.2.1 collection of data without consent which violates the privacy of individuals example, social media posts.
5.4.2.2 Inappropriate data sharing with third parties which could potentially harm the victims.
5.4.2.3 Re-identification of attacks which is a breach of confidentiality.

These risks are prevented through strict consent protocols, differential privacy to mitigate the chances of re-identification.
Machine Learning Models could also be trained through a federated learning system using decentralized data instead of direct raw data.

5.4.3 Informed Consent

There should be strict consent protocols before utilizing the Data of individuals. Such consent must be total and informed.

Most times, the participants don't understand how AI models use their data, so there’s a gap in the consent given. Furthermore, there coercion risks when patients in rehab are forced into the use of AI for treatment which is a form of exploitation of vulnerable persons.

Additionally, there are also instances where Data collected for depression related research collected from participants is then utilized for the prediction of addiction. Such actions extend behind the original consent given. As a result, there are potential situations in which the consent of the participant might be breached, or the data of the participant is used in a manner that extends beyond the initial consent given.
Considering these, Researchers must do well to mitigate the risks of a breach of consent.

To mitigate these risks, there should be:

5.4.3.1 Dynamic Consent Models, which allow patients to adjust permissions over time as well as observe how their data was used.
5.4.3.2 Re-Consent protocols when Data is to be used for some other research.

5.5 Real-World Application Scenarios

Artificial Intelligence could be used in workplaces, schools, etc. AI scan systems can detect risks of substance abuse before it gets worse. It can also help with:

- Personalized intervention for each patient without the risks of stigmatization. Basically, it can aid clinical decision support.
- Public Health Surveillance is another area where it could be applied, and the risks of mass substance abuse problems could be reduced.
- Artificial Intelligence provides the tools for Digital Therapeutics and Harm Reduction, through wearables and/or smartphones to track and predict addiction risks.
- AI Chatbots that help with personal and digital-driven therapy for individuals who shun traditional modes of therapy.

VI. CONCLUSIONS AND RECOMMENDATIONS.

6.1 Summary of the Research and its Contributions.

The importance of this research cannot be understated. Throughout the course of this work, the benefits of AI within Mental Health sectors were discussed.
Traditional Methods of Detecting Substance Abuse risk are usually not effective enough in such detection. Usually, it fails to detect Substance Abuse before it becomes worse. It usually detects Substance Abuse after the fact which is no good.
Artificial Intelligence and Machine Learning are capable of being trained in detecting Substance Abuse through behavioral Data and Digital Data trends.

The research also explored the challenges of using AI for Substance Abuse detection, such as data quality and bias, real world deployment etc.

6.2 Call to Action: Researchers, Policy Healthcare Providers, Policy Makers.

The actors involved in the important decisions that shape the research, important laws and policies as well as important medical advancements Mental Health treatment are important. As such, Researchers should continue to explore better ways of identifying Substance Abuse with aid of Artificial Intelligence. Researchers should avoid training Models with biases that prove detrimental to the research. Furthermore, Healthcare providers and the government should integrate healthcare with Artificial intelligence that helps with efficient and effective healthcare.

6.3 Future Work.

Future research should explore developments in important sectors of Artificial Intelligence and Mental Healthcare that creates better treatment and prevention of Substance Abuse. These areas include

6.3.1 Real-time Data
6.3.2 Wearable tech Integration.
6.3.3 Deployment in Clinical/community settings.

REFERENCES

[1] American Addiction Centers (2025). Alcohol and Drug Abuse Statistics (Facts about Addiction) https://americanaddictioncenters.org/rehab-guide/addiction-statistics-demographics

[2] Azzollina,D., Secchetin,E., Berchialla,P., Zhu,T., Acar, A.S., Gregori,D. (2025) Machine Learning Approaches for Monitoring Mental Health and Substance Abuse using Social Media Data. Frontiers in Public Health, 13, Article 1590564, https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2025.1590564/full

[3] Babu', K., Pilli., D., Sandeep, V., Deepthi, K. (2024) AI-Powered Interventions: Revolutionizing Drug Abuse Prevention. Journal of Drug and Alcohol Research, 13,Article 236687,1-5. https://doi.org /10.4303/jDAR/236287www.ashdin.com/articles/aipowered-interventions- revolutionizing-drug-abuse-prevention.pdf

[4] Bump, B.,(2025) Showcase 2025:UAlbany Researchers Leverage Machine Learning to Predict Depression in College Students. University of Albany. https://www.albany.edu/news- center/news/2025-showcase-2025-ualbany-researchers-leverage-machine-learning-predict

[5] Grosshans,M., Paul,T., Fischer,S.K.M,,Lotzmann,N., List,H., Haag,C., Jochen, M. (2024) Conversation-based AI for Anxiety Might Lower The Threshold for Traditional Medical Assistance: A Case Report.Frontiers in Public Health, 12, Article 1399702. https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2024.1399702/full

[6] Kim,S., Kim,H., Kim,S., Lee,H.,Ahmed Hammoodi A., Choi,Y., Kim, H.J., Smith,L., Kim,M.S.,

[7] Fond,G., Boyer,L.,Balk,S.W.,Lee,H.,Park,J., Kwon,R., Woo,S., Yon,D.K. (2025) Machine Learning-Based Prediction of Substance Use in Adolescents in Three Independent Worldwide Cohorts:Algorithm Development and Validation Study. JMIR Publications: Advancing Digital Health & Open Science, 27,Article e62805 https://www.jmir.org/2025/1/e62805

[8] Luhn,B., (2025) UH/UC Researchers Use Machine Learning To Map Brain Changes From Heroin Addiction. University of Houston. https://www.uh.edu/news- events/stories/2025/april/04302025-heroin-brain-study.php

[9] Mental Health and Lifestyle Habits (2019-2024) https://www.kaggle.com/datasets/atharvasoundankar/mental-health-and-lifestyle- habits-2019-2024

[10] Stadler, A., (2025) The 2025 State of Recovery: 5 trends in Behavioral Health You Need to Know About.Recovery.comhttps://www.ashdin.com/articles/aipowered-interventions- revolutionizing-drug-abuse-prevention.pdf

[11] Tan, M., Hatef, E., Taghipour,D., Vyas,K., Kharrazi,H.,Gottlieb,L., Weiner, J. (2020) Including Social and Behavioral Determinants in Predictive Models: Trends, Challenges and Opportunities. JMIR Publications:Advancing Digital Health and Open Science 8(9), https://medinform.jmir.org/2020/9/e18084/

[12] World Health Organization, Africa Region. Substance Abuse(2025) https://www.afro.who.int/health-topics/substance-abuse