ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/378430682

Towards Data-Driven Artificial Intelligence Models for Monitoring, Modelling and Predicting Illicit Substance Use

Conference Paper · January 2024

DOI: 10.1007/978-3-031-54820-8_29

CITATION
1

READS
311

7 authors, including:

Elliot Mbunge
University of Johannesburg
90 PUBLICATIONS **2,420** CITATIONS
SEE PROFILE

Benhildah Muchemwa
University of Eswatini
17 PUBLICATIONS **675** CITATIONS
SEE PROFILE

John Batani
Botho University
38 PUBLICATIONS **847** CITATIONS
SEE PROFILE

Itai Chitungo
University of Zimbabwe
48 PUBLICATIONS **1,345** CITATIONS
SEE PROFILE

All content following this page was uploaded by Enos Moyo on 21 May 2024.

The user has requested enhancement of the downloaded file.

Towards Data-Driven Artificial Intelligence Models for Monitoring, Modelling and Predicting Illicit Substance Use

Elliot Mbunge[1,2](✉), John Batani[3], Itai Chitungo[4], Enos Moyo[5], Godfrey Musuka[6], Benhildah Muchemwa[2], and Tafadzwa Dzinamarira[7]

[1] Department of Information Technology, Faculty of Accounting and Informatics, Durban University of Technology, P O Box 1334, Durban 4000, South Africa
mbungeelliot@gmail.com
[2] Department of Computer Science, Faculty of Science and Engineering, The University of Eswatini (Formerly Swaziland), Kwaluseni, Manzini, Eswatini
[3] Faculty of Engineering and Technology, Botho University, Maseru 100, Lesotho
[4] Department of Medical Laboratory Sciences, University of Zimbabwe, Harare, Zimbabwe
[5] Medical Centre Oshakati, Oshakati, Namibia
[6] ICAP at Columbia University, Harare, Zimbabwe
[7] School of Health Systems and Public Health, University of Pretoria, Pretoria, South Africa

Abstract. Illicit substance use (ISU) is a major public health problem and a significant cause of morbidity and mortality globally. Early assessment of risk behaviour, predicting, identifying risk factors, and detecting illicit substance use become imperative to reduce the burden. Unfortunately, current digital tools for early detection and modelling ISU are largely ineffective and sometimes inaccessible. Data-driven artificial intelligence (AI) models can assist in alleviating the burden and tackling illicit substance use but their adoption and use remain nascent. This study applied the PRISMA model to conduct a systematic literature review on the application of artificial intelligence models to tackle illicit substance use. The study revealed that elastic net, artificial neural networks support vector machines, random forest, logistic regression, KNN, decision trees and deep learning models have been used to predict illicit substance use. These models were applied to tackle different substance classes, including alcohol, cannabis, hallucinogens, tobacco, opioids, sedatives, and hypnotics among others. The models were trained and tested using various substance use data from social media platforms and risk factors such as socioeconomic and demographic data, behavioural, phenotypic characteristics, and psychopathology data. Understanding the impact of these risk factors can assist policymakers and health workers in effective screening, assessing risk behaviours and, most importantly, predicting illicit substance use. Using AI models and risk factors to develop data-driven intelligent applications for monitoring, modelling, and predicting illicit substance use can expedite the early implementation of interventions to reduce the associated adverse consequences.

Keywords: Artificial Intelligence · Illicit Substance Use · Data-driven · Africa

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
R. Silhavy and P. Silhavy (Eds.): CoMeSySo 2023, LNNS 935, pp. 361–379, 2024.
https://doi.org/10.1007/978-3-031-54820-8_29

1 Introduction

Illicit substance use remains a major global public health problem, and the outbreak of coronavirus disease 2019 (COVID-19) exacerbated the burden [1]. Drug use disorders caused 85,984 deaths worldwide in 2019 (55,616 deaths in males and 30,367 deaths in women), accounting for 47% of all global deaths caused by drug-related disorders [1]. In addition, the outbreak of COVID-19 and imposed movement restrictions increased the burden which consequently has psychological effects such as stress, anxiety, depression, and other mental health issues exacerbated by socioeconomic challenges. These psychological effects played a significant role in the growth of illicit substance use (ISU) incidences [2]. For instance, people aged between 15 and 64 years, were among 284 million illicit drug users globally in 2020, a 26% increase from the previous decade [1]. Most patients in Africa and South America receiving treatment for ISU problems are under 35 years of age [1]. The use of illicit substances has increased substantially among adolescents, and they tend to develop addiction because of the propensity for experimenting, curiosity, low self-esteem and lack of guidance and counselling services. This is also attributed to early mental and behavioural health issues, peer pressure, a bad family structure, inadequate parental supervision and interactions, lack of opportunities, loneliness, and easy access to drugs [3].

Moreso, protective measures such as peer groups that stimulate high self-esteem, religiosity, peer influence, self-control, parental supervision, academic proficiency, rehabilitation, anti-drug use legislation, and a strong sense of community connectedness have been utilized to tackle and reduce the catastrophic impact of ISU [4]. Preventing ISU can reduce its catastrophic impact on health, as it affects youths to properly transition into adulthood by impeding the development of critical thinking and learning skills [5]. In addition, the use of illicit substances also affects sexual and reproductive health (SRH) [6] and poor reproductive outcomes, such as unintended pregnancies, preterm deliveries, and maternal and neonatal morbidity and mortality [7]. ISU can also result in chronic illnesses [8], and socioeconomic problems such as violence, dropping out of school, family breakup, unemployment, and living in squalor and poverty [9]. Due to the severity of the consequences associated with ISU, it is imperative to develop AI models that can predict, detect, monitor, and model risk factors associated with ISU to assist healthcare professionals in designing effective interventions to reduce the burden.

1.1 Significance of the Study

ISU can be predicted using interviewer-administered questions or self-administered methods, but they need extra personnel and time. As a result, few health systems routinely screen for ISU [10]. Patients can be screened for ISU using an automated technique that uses information gathered during routine health treatment. Therefore, artificial intelligence (AI) models can be used to monitor and predict ISU among different populations. AI models can learn to carry out specific tasks by identifying patterns in data [11]. By concentrating on linking input features like psychological, physical, and environmental factors with ISU, it stands as a potent alternative to data-driven models for identifying ISU vulnerability and predicting illicit substance use [12]. Several studies focused on the social, psychological, and mental consequences of illicit substance use [3, 13–15], while some applied statistical tools to survey data to determine risk factors associated with ISU [7] and recommend protective measures [16]. In addition to survey data on ISU, there is positive progress, though at a slower pace, in utilizing smart wearable technologies such as smartwatches and wristbands to collect data on substance use [17, 18]. Such massive data can be utilized to develop data-driven intelligent models to tackle illicit substance use. Despite this, there is a dearth of literature on the application of AI models for monitoring, modelling and predicting illicit substance use. Such data-driven AI models can tremendously assist healthcare professionals in predicting the possibility of substance abuse, remote monitoring patients in rehabilitation and most importantly, modelling risk factors associated with ISU. This can also assist policymakers in developing and implementing evidence-based frameworks and guidelines for integrating AI into health systems to tackle ISU. Such applications are rare, but they can guide the future development of digital health interventions and tools for reducing ISU [19]. Therefore, this study presents a systematic literature review of AI models, risk factors associated with ISU and barriers and challenges hindering the development and integration of data-driven AI models into health information systems. The study also presents recommendations to address the identified barriers and challenges.

The remainder of this paper is structured as follows. Section 2 presents the methodology used in conducting the systematic review. The results analysis and discussion of AI models and applications are presented in Sect. 3. The barriers, challenges and recommendations for integrating data-driven artificial intelligence models for tackling ISU are presented in Sect. 4, while Sect. 5 presents the conclusion.

2 Materials and Method

2.1 Study Design

The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) model guided this systematic review in which primary and scientific articles that were published between January 2015 and June 2023 on substance abuse were retrieved and analysed. The study was conducted between 16 February and 14 June 2023. This study sought to identify AI models used in substance abuse research, their purposes, performance and limitations and the risk factors used. Figure 1 summarises the search results, study selection and inclusion process followed in this study.

2.2 Search Strategy

The major databases searched were Web of Science, Google Scholar, Scopus and Science Direct, using the keywords "*artificial intelligence*", "*deep learning*", "*machine learning*", "*illicit substance abuse*", and "*substance use disorders*". The publication year was restricted to between 1 January 2015 and 14 June 2023. Additional articles were identified through citation chaining.

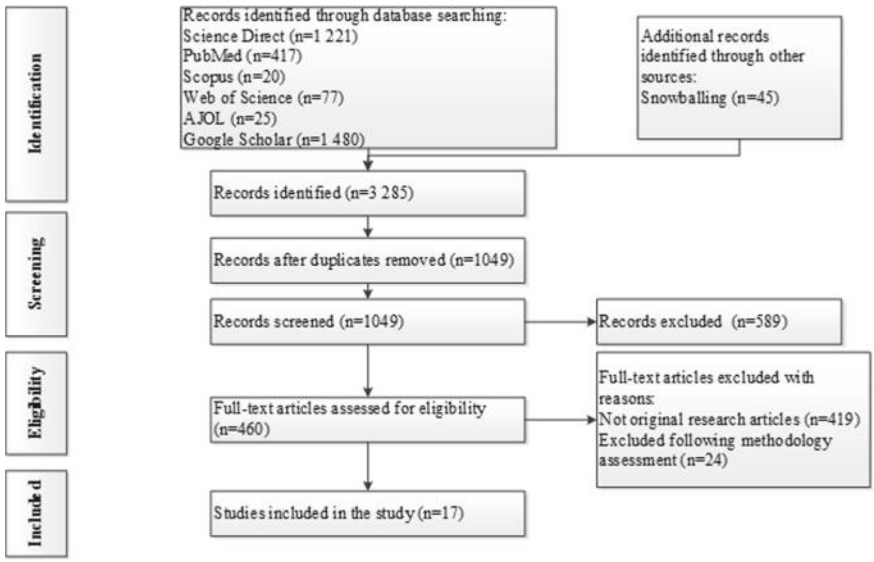

Fig. 1. PRISMA model

2.3 Study Selection and Eligibility Criteria

Articles returned from database searches were initially screened based on the abstracts, and where a conclusive decision could not be made on eligibility, the full articles were retrieved and screened. Articles that were deemed eligible based on screening were assessed for eligibility by methodology assessment and article type. Eligible articles were those that were published between 1 January 2015 and 14 June 2023 when the study was concluded. The eligibility criteria were thus: articles that had low perceived risk-of-bias, were written in English, or had English translations, peer-reviewed, empirical and applied any type of AI algorithm to predict, detect, assess or examine substance abuse, including determining the risk factors. Any articles that did not meet the specified criteria, were deemed to have a high risk of bias, or had poor methodologies were excluded.

2.4 Data Extraction

Two researchers independently screened all articles based on titles, abstracts, and quality of methodology. The researchers resolved any differing views on papers by conducting a meeting in which each explained their reasoning in line with the prescribed criteria until a consensus was reached. The two authors involved in article retrieval, screening and assessment continuously engaged each other throughout the data extraction process. Data were extracted into a standard table, and both researchers checked the data for accuracy.

2.5 Risk-of-Bias and Quality Assessment of Included Sources

The included articles were assessed for quality and risk of bias using the Critical Appraisal Skills Programme checklist [20]. The assessment criteria were methodology appropriateness and validity of findings, facilitated by the Critical Appraisal Skills Programme questionnaire, helping eliminate systematic biases [21]. Conducting and visualizing the results of this process was enabled by using the Review Manager software called RevMan5.4.1 (Table 1).

Table 1. Application of AI models to tackle illicit substance use

Reference	Model(s)	Purpose	Risk Factors	Performance	Limitations
[12]	Naïve Bayes, SVM and Random Forest	Evaluation of Substance use disorder	Phenotypic characteristics and environmental factors	74% and 86% accuracy for different age groups	ML algorithms delineate the psychological, health, and environmental characteristics associated with the risk for SUD
[22]	Random Forest	Predicting the severity of substance use	Behavioural characteristics	71% and 91% accuracy for different age groups	The harmfulness score does not fully account for cumulative exposure to each substance A longitudinal study is required to measure substance use severity
[23]	Logistic regression	Predicting ISU	Environmental factors, family substance use,		Limited sample size and survey data are subject to biases
[24]	XGBoost	Predicting Susceptibility to substance abuse	Behavioural characteristics	ROC:99% PRC:98%	Drug interactions involving opioids are solely determined by how frequently they are prescribed and when side effects occur

(*continued*)

Table 1. (*continued*)

Reference	Model(s)	Purpose	Risk Factors	Performance	Limitations
[25]	Negation detection algorithm, NegEx., NLP	Detection of Substance-Use Status	Frequency, Type, Amount of smoking, alcohol, or drugs consumed by the patient and quit time and period	F1-score -99%	The dataset used was small. A larger dataset may enhance the capacity to produce additional rules
[26]	SVM	Predict substance abuse treatment outcomes	Clinical data such as cocaine, methamphetamine, or heroin-dependent at the time of incarceration, no history of head injury and no history of psychosis	Accuracy: 80.58% Sensitivity: 81.31% Specificity: 78.13%	The negative predictive value is lower than desired and misidentifies someone with a higher risk for discontinuation
[27]	Elastic net	Assess ISU predictors	characteristics and psychopathology data	AUC: 83%	Instead of conducting a formal clinical interview, the screening test (ASSIST) was utilised to determine whether a person used illicit drugs
[28]	Decision trees, SVM and Boosted decision trees	Detecting ISU Recovery Problems	Peer-to-peer discussion forums and ISU self-management pages on the forum	F1-scores DT:88% SVM:89% Boosted DT:94%	The model did not label subtypes of recovery problems. Participants who did not post on the forum were excluded
[29]	SVM, Random forest, Naïve Bayes and Logistic regression	Substance Abuse Risk Assessment	Social media data	F1 scores SVM:82.6% RF: 85.9%, Naïve Bayes: 85.3% LR: 88.2%	High data imbalance affects biomarker prediction. Data were collected from limited social media platforms

(*continued*)

Table 1. (*continued*)

Reference	Model(s)	Purpose	Risk Factors	Performance	Limitations
[30]	RF, KNN, DT, Linear SVC, Gaussian Naïve Bayes and Logistic Regression	Predicting Individual Substance Abuse Vulnerability	Demographic and Behavioural characteristics	Accuracy RF: 95.08%, KNN: 88.52%, DT: 85.24%, Linear SVC: 95.90%, Gaussian Naïve Bayes: 92.62% and LR: 94.26%	Data were collected from the same location and same age group which might affect the performance of the models when exposed to new data from other locations
[31]	Random forest, Super-learning and Artificial neural networks	Predict the efficacy of treatment for substance use disorders	Patient characteristics, treatment characteristics and type of problematic substance	AUC for the models ranged between 79.3% and 82%	There are relatively few details about the kind of treatment and how it was delivered
[32]	Random forest	Determine the socioeconomic causes of patients abandoning substance abuse treatment	Demographic and Behavioural characteristics	AUC of 89%	Not specified
[33]	Artificial Neural networks (ANN)	Predict volatile substance abuse for drug risk analysis	Agreeableness, conscientiousness extraversion, neuroticism, openness to experience, impulsiveness, sensation seeking and demographic details	Accuracy of 81.1%	The study used an open-source dataset; therefore, the model needs to be validated with real data
[34]	Natural language processing	Identification of Substance Abuse	Status, type, method, amount, frequency, exposure history and quit history	F1-score between 80%–91%	Not mentioned

(*continued*)

Table 1. (*continued*)

Reference	Model(s)	Purpose	Risk Factors	Performance	Limitations
[35]	Machine learning regression algorithm (actual algorithm not specified in the paper)	Prediction of severity of alcohol use disorders	Neural features using imaging data and self-reporting	The model correctly explained 33% of the variance	They used multimodal features hoping their model would perform better than those that used unimodal features, but this hypothesis was not supported by their results
[36]	Logistic regression, Naïve Bayes and gradient boost	Prediction of alcohol use	Not clear from the paper	Accuracy of 97.55%	Not mentioned
[37]	Random forest	Predicting comorbid substance use disorders among people with bipolar disorder	Sociodemographic and clinical data	F1-score:66% Accuracy:65.3% Sensitivity:69.6% Specificity:61.2%	Clinical variables collected retrospectively from electronic clinical records may have affected the accuracy and reliability of the data

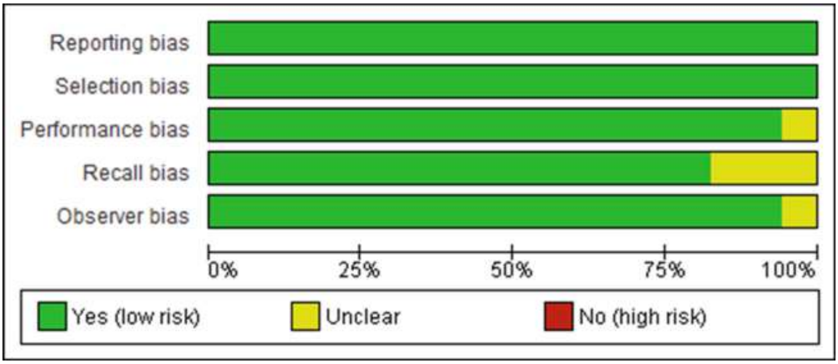

Fig. 2. Summary of RoB assessment for all included studies

3 Results Analysis and Discussion

3.1 Analysis of Risk-of-Bias

The researchers analysed the risk-of-bias of the included papers, focusing on reporting, selection, performance, recall and observer biases. Figure 2 presents the summarized analysis for all the included articles.

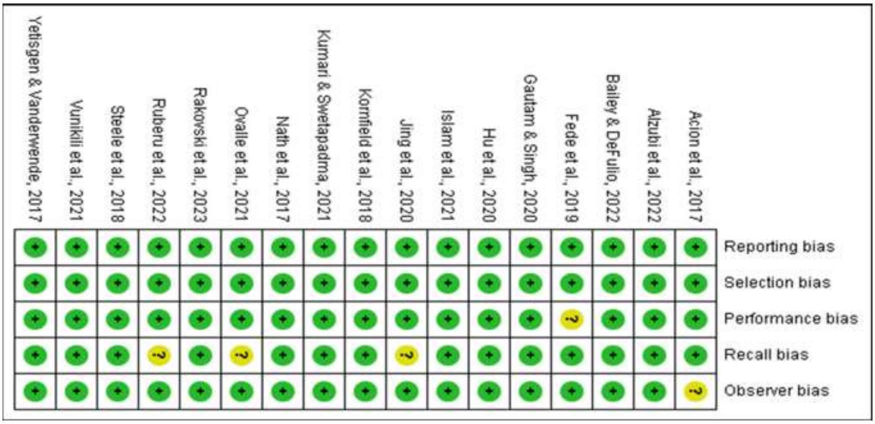

Fig. 3. Analysis of RoB for individual studies.

Key: low risk unclear

Figure 3 shows the analysis of bias for each included paper. The researchers' perceptions on the risk-of-bias of the included papers were that there was low bias across all the five bias types. The following sections present an analysis of the findings regarding AI models and risk factors.

3.2 Artificial Intelligence Models for Predicting Illicit Substance Abuse

Artificial intelligence focuses on developing intelligent models (algorithms) and smart machines that can perform tasks that typically require human intelligence [38]. It has various subsets, including deep learning, natural language processing and machine learning. Machine learning (ML) is a subset of AI that has been significantly used in healthcare to perform various tasks, including diagnosis [39], detection [40], prediction [41, 42] and classification. ML models learn from samples to solve classification, association and clustering problems in healthcare [43]. This study has revealed that several AI models have been used in predicting, modelling and monitoring substance use, all of which fall under machine and deep learning. Deep learning is a subset of ML that has more layers than a neural network and automates feature extraction, removing the hand-tuning that is predominant in traditional ML. There has been significant progress in detecting substance use disorder [44], assessing future risk and predicting treatment success through the use of machine learning models. For instance, a study conducted by [45] applied machine learning models such as decision tree (DT), random forest (RF), logistic regression (LR), K-Nearest Neighbours (KNN), Gaussian Naïve Bayes and Support Vector Machine (SVM) to predict individual substance abuse. Figure 4 is a word cloud for the identified AI models and algorithms used in identifying and predicting substance abuse.

Fig. 4. Word cloud for a few identified models for tackling ISU

SVM is a widely used technique that defines an optimal hyperplane to distinguish between items falling in classes of interest [41]. It is an ML algorithm used to solve classification and regression problems using labelled data [41]. A Cartesian plane is normally used to illustrate SVM, where a hyperplane is drawn to separate observations in different classes or categories [41]. Observations are represented as coordinates on the Cartesian plane, and these coordinates are called support vectors. Islam et al. [45] used the SVM algorithm to identify vulnerabilities to substance abuse.

Random forest is an ML algorithm that uses labelled data to solve classification or regression problems [32, 42]. The algorithm combines multiple decision trees to determine the output, thus, it is ensemble-based [39, 42, 46]. Each tree in the forest produces the output that it uses to vote for the output, and the output that gets the majority vote becomes the forest's (model's) overall output [42]. The advantages of this algorithm include high accuracy and efficiency, as well as the ability to ingest datasets that contain both numerical and categorical predictors [42]. Islam et al. [45] created a random forest model to identify vulnerabilities to substance abuse, while [47] used it to predict substance use disorder treatment outcomes.

Transfer learning entails "repurposing a trained model on a related task" [47] and is mainly used in deep learning to leverage related pre-trained models where datasets are small. Bailey and DeFulio [47] used transfer learning (and compared it with a random forest model) to predict substance use disorder treatment outcomes.

Logistic regression is a machine learning algorithm that uses labelled data to solve classification problems [41, 42, 48]. It uses the sigmoid activation function to return a value (0 or 1) that signifies whether a given observation belongs to a given class or not, using the formula below [41].

$$f(z) = (1 + e^{-z})^{-1} \tag{1}$$

where z is "the weighted sum of each neuron" [41].

Since the sigmoid function's output must either be 0 or 1, any continuous value obtained is set to 0 or 1 based on which value it is closer to. Thus, a threshold value, normally 0.5, is used. Logistic regression was used in creating a predictive classifier for identifying vulnerabilities to substance abuse [45], alcohol use [36] and substance use disorders [31].

Naïve Bayes is a supervised classification machine learning algorithm based on the Bayes theorem [49]; thus, it is probabilistic. The Gaussian Naïve Bayes, like the Naïve Bayes, is a probabilistic classification algorithm based on the Gaussian distribution [50]. The algorithm assumes that each feature is independently capable of predicting the target variable [50]. Naïve Bayes has been used to predict alcohol use [36], while Islam et al. [45] used Gaussian Naïve Bayes to identify vulnerabilities to alcohol use.

Super learning is an ensemble-based ML algorithm that was derived from the stacking algorithm, whose output is the weighted mean of all the included predicted algorithms [31]. Also, a study conducted by [31] used superlearning to predict SUD treatment, with regression, RF, and deep neural networks as the constituent algorithms. They used the area under the curve as an evaluation metric and it ranged between 79.3% and 82.0% as reported in their study.

Though a decision tree is a simple ML algorithm, it is widely used to solve classification problems [41]. The algorithm uses a rooted-tree data structure with a root, leaves and internal nodes, where internal nodes lie between the root and terminal (leaf) nodes [51]. Classes or labels in a decision tree are represented by terminal nodes (leaves), while test conditions (attributes/ characteristics) are represented as inner and root nodes [42]. A decision tree segments the feature space into several simple regions. Boosted decision trees are a type of decision tree that utilises the ensemble approach, in which each tree learns from the residual of the trees, enhancing the overall performance.

Islam et al.[45] used a decision tree to identify vulnerabilities to substance abuse. Gradient boost was used by Kumari & Swetapadma [36] to predict alcohol use and produced close to 98 per cent accuracy. Moreover, a study conducted by Kornfield et al. [28] applied SVM, decision trees and boosted decision trees to detect illicit substance use using a natural language processing program called Linguistic Inquiry and Word Count and data from online substance abuse forums. Their study revealed that boosted decision trees outperformed other models with an F1 score of 94%. However, participants who did not post on the forum were excluded, and the model could not label subtypes of recovery problems. A study by Rakovski et al. [27] applied an elastic net to assess the performance of the selected illicit substance abuse predictors. The model achieved the highest area under the receiver operating characteristics curve (AUC) of 83%. The lasso and ridge regression regularisations are combined linearly by the elastic net approach to creating models that can remove unnecessary variables while maintaining small coefficients, improving generalisation.

Natural language processing algorithms have been used to detect substance abuse from clinical textual data. For instance, a study conducted by Alzubi et al. [25] applied a negation detection algorithm called NegEx to detect substance use status and the algorithm achieved an F1-score of 99%. Also, Yetisgen & Vanderwende [34] applied natural language processing to the Identification of substance abuse from social history in clinical text automatically and achieved FI-score ranging between 80% and 91%. Interestingly, social media platforms together with natural language processing algorithms have been used to detect, identify and assess substance abuse risk behaviour. For instance, a study by Ovalle et al. [29] utilised social media data and machine learning models such as LR, linear SVM, Naïve Bayes, and RF to assess substance abuse risk behaviour among men who have sex with men(MSM). Such modes tremendously assist in implementing adaptive interventions to target substance use risk behaviour among hard-to-reach groups such as MSM. These algorithms were used for different purposes in substance abuse research. Figure 5 shows the different purposes for which the explained AI algorithms were used, and these are presented in the form of a word cloud.

Fig. 5. Application of AI models to tackle ISU.

3.3 Risk Factors for Modelling Substance Abuse Using Artificial Intelligence Models

Illicit substance use is associated with various factors, including psychological, and socioeconomic data, sociodemographic data, phenotypic characteristics and environmental factors, family setup, and religion. A study conducted by [12] applied random forest, SVM, MLP, and logistic regression to analyse substance use and its outcomes using psychological self-regulation spanning behaviour control data and daily routine data. Additionally, Rakovski et al. [27] utilised socioeconomic and sociodemographic data as predictors for predicting ISU among young adults. Predictors such as age, gender, religion, alcohol dependence, tobacco abuse, and family history of alcohol dependence were significantly associated with illicit substance use. Also, a study by Ruberu et al. [23] utilised sociodemographic data (culture, gender, age), family substance use and general environmental factors such as early life stress, peer friends, level of parental monitoring and involvement of parents in daily activities. These factors were used to predict illicit substance use among adolescents using multivariate covariance generalised linear models. Age, early life stress, lifetime substance use, age of first use, maternal education, parental attachment, family cigarette use, and family history of substance use are some risk factors that have been used to predict the use of illicit substance use [27]. Conversely, a systematic review conducted by Nawi et al.[3] classified ISU risk factors into three main groups; community, individual and family factors. Community factors include having friends and peers who abuse substances and the community's culture and beliefs. Individual factors include low religion, peer pressure, negative upbringing, psychiatric disorders, rebelliousness, exposure to hazardous substances and behavioural addiction, and easy and high accessibility to illicit substances. Family risk factors consist of a family history of substance use, maternal smoking, the family's abusive and addictive behaviour, poor level of monitoring, and negligence [3]. These factors are paramount in predicting illicit substance use.

4 Barriers, Challenges and Recommendations for Integrating Data-Driven Artificial Intelligence Models for Tackling Illicit Substance Use

Applying artificial intelligence techniques such as deep learning and machine learning models for modelling risks and predicting illicit substance use requires massive data for training, testing, and sometimes validating the performance of the model. However, applying such models encounters impediments as illicit substance use data generally becomes not readily available [52]. Several studies applied artificial intelligence models on survey data to predict the severity of substance use [22] and the evaluation of substance use disorder [12]. A study by Walsh et al.[52] highlighted that the shortage of data is caused by numerous factors, including stigma, patients not seeking treatment, self-discriminating and stereotyping behaviour, reduced independence for patients abusing substances and criminalisation of ISU.

Artificial intelligence algorithm bias has been reported by Walsh et al. [52] as one of the major barriers to integrating data-driven artificial intelligence-based applications for tackling illicit substance use. Having a biased artificial intelligence algorithm means that key connections between the input features and the output variable are being missed which consequently affects the performance of the application. Reducing bias tends to increase the performance of the model; however, trade-offs between performance, underfitting and overfitting should be carefully considered and evaluated. Bias can happen in data, in model specification [52], in training, testing, validating and deployment of artificial intelligence algorithms, especially in machine learning. Therefore, a robust discussion should address the necessity of data reuse and sharing to increase algorithm transparency and improve algorithm correctness and reliability.

AI-based applications require consistently measured timely data to make long-term predictions. However, constraints due to structural barriers exist that limit access to quality data. Constraints such as enabling policies, poor reporting of substance abuse [53] or lack of electronic health records. Electronic health records (EHR) have been demonstrated to be a crucial tool for enhancing patient information access and the standard of care. EHR hasn't, however, been widely used or adopted in SSA [54, 55]. Several countries in SSA lack a clear policy on the implementation of EHRs as well as any financial incentives to steer the adoption of EHRs. In SSA countries, there are competing needs for already challenged health systems such that substance abuse is not a priority. This impacts the ability of available data for modelling, monitoring and predicting substance abuse due to missing data points. Developing AI modelling with missing data may lead to bias and loss of precision ('inefficiency') [56]. Secondary data from drug market surveillance, drug testing services and wastewater is also useful in providing the often missing part of the equation. Limiting factors for most of these secondary data sources are restricted access to public health [53].To improve access to timely data on substance abuse, there is a need to implement epidemiological surveillance infrastructure, which requires the removal of structural barriers. Countries must be urged to invest in or improve health systems that promote the collection and sharing of data on substance abuse.

Once the data constraint hurdle is overcome, the next challenge is identifying the most appropriate indicators of illicit substance use severity. There is no clear consensus on a definition of severity and is usually inferred from the frequency of consumption or symptoms because of ISU. These metrics of severity are subject to and lead to different interpretations making it difficult to standardise the indicators of ISU severity [22]. The combination of missing data points and choice of indicators introduces biases in algorithms. The biases include "patients not identified by algorithms, sample size and underestimation, and misclassification and measurement error." These biases result from socioeconomic disparities in health care [57].

Several studies reported the lack of rehabilitation facilities in many SSA countries. The shortages of rehabilitation facilities could be alleviated by adopting and implementing digital rehabilitation services. Though this is infancy and more research is needed, digital rehabilitation services are delivered through digital tools such as smart applications focused on improving patient outcomes and relapse prevention [58]. A digital rehab start-up Workit Health offers group therapy, coaching, and medication-assisted treatment through a mobile application.

Insufficient health workers, especially psychiatrists have been reported in many studies of SSA. Addressing the shortages in the mental health workforce requires more than scaling up the training of psychiatrists, psychologists and psychiatric nurses, to task shifting-delegating healthcare tasks from specialists to various non-specialist health professionals and other health workers [59]. Additionally, integration of mental health services at primary health care is delivered through community-based and task-sharing approaches.

The competing needs of health systems plus the stigma of ISU make it challenging to provide adequate health care to the user of drug abuse. Drawing from the emerging digital phenotyping approaches which leverage ubiquitous sensing technology [60], AI data-driven smart applications can be developed for remote monitoring, tracking and reporting those with substance abuse tendencies. However, implementation and adoption of these emerging technologies require appropriate funding of digital infrastructure and m-Health policies. Most countries especially in LMICs lack political will and digital health governance policies and this requires urgent redress through crafting policies and frameworks that support the use of AI in the fight against SUB [61].

5 Conclusion

Illicit substance use has increased substantially globally, and early detection, prediction and identification of risk groups is imperative to reduce its impact. Due to the dearth of literature in this emerging research area, we conducted a systematic review on the application of AI models for monitoring, modelling and predicting illicit substance use. The study revealed that there is positive progress towards the application of artificial intelligence algorithms to tackle illicit substance abuse. AI models have been used to perform various tasks, including predicting, identifying high-risk groups, detecting ISUs, and assessing ISU predictors to develop and expedite the early implementation of interventions that reduce the associated adverse consequences. Age, gender, early life stress, lifetime substance use, maternal education, parental attachment, having friends who abuse drugs, culture and religion, family cigarette use, and family history of substance use are some risk factors that have been used to predict the use of illicit substance use using artificial intelligence algorithms. Support vector machine, random forest, logistic regression, KNN, decision trees and natural language processing algorithms are among AI algorithms that have been predominantly used to analyse and predict illicit substance use. Such models can tremendously assist in the identification of ISU risk factors among youths, adolescents, and adults. These models can be used further to develop data-driven artificial intelligence-based ISU tools to identify individuals at risk and alert health workers to provide appropriate interventions and prevention measures.

References

1. WHO. World Drug Report 2022. WHO 2022. https://www.unodc.org/unodc/data-and-analysis/world-drug-report-2022.html. Accessed 9 Apr 2023
2. Mukwenha, S., Murewanhema, G., Madziva, R., Dzinamarira, T., Herrera, H., Musuka, G.: Increased illicit substance use among Zimbabwean adolescents and youths during the COVID-19 era: an impending public health disaster. Addiction **117**, 1177–1178 (2022). https://doi.org/10.1111/ADD.15729
3. Nawi, A.M., et al.: Risk and protective factors of drug abuse among adolescents: a systematic review. BMC Publ. Health **21**, 1–15 (2021). https://doi.org/10.1186/S12889-021-11906-2/FIGURES/2
4. Drabble, L., Trocki, K.F., Klinger, J.L.: Religiosity as a protective factor for hazardous drinking and drug use among sexual minority and heterosexual women: findings from the national alcohol survey. Drug Alcohol Depend. **161**, 127–134 (2016). https://doi.org/10.1016/J.DRUGALCDEP.2016.01.022
5. Crews, F., He, J., Hodge, C.: Adolescent cortical development: a critical period of vulnerability for addiction. Pharmacol. Biochem. Behav. **86**, 189–199 (2007). https://doi.org/10.1016/J.PBB.2006.12.001
6. Hall, K.S., Moreau, C., Trussell, J.: The link between substance use and reproductive health service utilization among young US women. Subst. Abuse **34**, 283–91 (2013). https://doi.org/10.1080/08897077.2013.772934
7. Reardon, D.C., Coleman, P.K., Cougle, J.R.: Substance use associated with unintended pregnancy outcomes in the national longitudinal survey of youth **30**, 369–83 (2004). https://doi.org/10.1081/ADA-120037383

8. Fergie, L., Campbell, K.A., Coleman-Haynes, T., Ussher, M., Cooper, S., Coleman, T.: Identifying effective behavior change techniques for alcohol and illicit substance use during pregnancy: a systematic review. Ann. Behav. Med. **53**, 769–781 (2019). https://doi.org/10.1093/abm/kay085
9. Nyaga, J.: Socio-economic and health consequences of drugs and substance use in Gachie, a Peri-urban town on the outskirts of Nairobi. Afr. J. Alcohol Drug Abuse **6** (2021)
10. Afshar, M., et al.: Development and multimodal validation of a substance misuse algorithm for referral to treatment using artificial intelligence (SMART-AI): a retrospective deep learning study. Lancet Digit. Health **4**, e426–e435 (2022). https://doi.org/10.1016/S2589-7500(22)00041-3
11. Mbunge, E., Batani, J.: Application of deep learning and machine learning models to improve healthcare in sub-Saharan Africa: emerging opportunities, trends and implications. Telemat. Inform. Rep. **11**, 100097 (2023). https://doi.org/10.1016/J.TELER.2023.100097
12. Jing, Y., et al.: Analysis of substance use and its outcomes by machine learning I. Childhood evaluation of liability to substance use disorder. Drug Alcohol Depend **206**, 107605 (2020). https://doi.org/10.1016/J.DRUGALCDEP.2019.107605
13. Semple, D.M., McIntosh, A.M., Lawrie, S.M.: Cannabis as a risk factor for psychosis: systematic review **19**, 187–94 (2005). https://doi.org/10.1177/0269881105049040
14. Guxensa, M., Nebot, M., Ariza, C., Ochoa, D.: Factors associated with the onset of cannabis use: a systematic review of cohort studies. Gac. Sanit. **21**, 252–260 (2007). https://doi.org/10.1157/13106811
15. Moore, T.H., et al.: Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. Lancet **370**, 319–328 (2007). https://doi.org/10.1016/S0140-6736(07)61162-3
16. Nargiso, J.E., Ballard, E.L., Skeer, M.R.: A systematic review of risk and protective factors associated with nonmedical use of prescription drugs among youth in the United States: a social ecological perspective **76**, 5–20 (2015). https://doi.org/10.15288/JSAD.2015.76.5
17. Mahmud, M.S., Fang, H., Carreiro, S., Wang, H., Boyer, E.W.: Wearables technology for drug abuse detection: a survey of recent advancement. Smart Health **13**, 100062 (2019). https://doi.org/10.1016/J.SMHL.2018.09.002
18. Kunchay, S., Abdullah, S.: WatchOver: using apple watches to assess and predict substance co-use in young adults. UbiComp/ISWC 2020 Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers, pp. 488–93 (2020). https://doi.org/10.1145/3410530.3414373
19. Hamideh, D., Nebeker, C.: The digital health landscape in addiction and substance use research: will digital health exacerbate or mitigate health inequities in vulnerable populations? Curr. Addict. Rep. **7**, 317–332 (2020). https://doi.org/10.1007/S40429-020-00325-9/TABLES/3
20. CASP-UK. CASP CHECKLISTS. CASP-UK Website (2021)
21. Batani, J., Maharaj, M.S.: Towards data-driven models for diverging emerging technologies for maternal, neonatal and child health services in sub-Saharan Africa: a systematic review. Glob. Health J. (2022). https://doi.org/10.1016/J.GLOHJ.2022.11.003
22. Hu, Z., et al.: Analysis of substance use and its outcomes by machine learning: II. Derivation and prediction of the trajectory of substance use severity. Drug Alcohol Depend. **206**, 107604 (2020). https://doi.org/10.1016/J.DRUGALCDEP.2019.107604
23. Ruberu, T.L.M., et al.: Joint risk prediction for hazardous use of alcohol, cannabis, and tobacco among adolescents: a preliminary study using statistical and machine learning. Prev. Med. Rep. **25**, 101674 (2022). https://doi.org/10.1016/J.PMEDR.2021.101674

24. Vunikili, R., Glicksberg, B.S., Johnson, K.W., Dudley, J.T., Subramanian, L., Shameer, K.: Predictive modelling of susceptibility to substance abuse, mortality and drug-drug interactions in opioid patients. Front. Artif. Intell. **4**, 172 (2021). https://doi.org/10.3389/FRAI.2021.742723/BIBTEX
25. Alzubi, R., Alzoubi, H., Katsigiannis, S., West, D., Ramzan, N.: Automated detection of substance-use status and related information from clinical text. Sensors **22**, 9609 (2022). https://doi.org/10.3390/S22249609
26. Steele, V.R., et al.: Machine learning of functional magnetic resonance imaging network connectivity predicts substance abuse treatment completion. Biol. Psychiatry Cogn. Neurosci. Neuroimaging **3**, 141–149 (2018). https://doi.org/10.1016/J.BPSC.2017.07.003
27. Rakovski, C., et al.: Predictors of illicit substance abuse/dependence during young adulthood: a machine learning approach. J. Psychiatry Res. **157**, 168–73 (2023). https://doi.org/10.1016/J.JPSYCHIRES.2022.11.030
28. Kornfield, R., et al.: Detecting recovery problems just in time: application of automated linguistic analysis and supervised machine learning to an online substance abuse forum. J. Med. Internet Res. **20**(6), e10136 (2018). https://doi.org/10.2196/10136
29. Ovalle, A., et al.: Leveraging social media activity and machine learning for HIV and substance abuse risk assessment: development and validation study. J Med Internet Res **23**(4), e22042 (2021). https://doi.org/10.2196/22042
30. Islam, U.I., Sarker, I.H., Haque, E., Hoque, M.M.: Predicting individual substance abuse vulnerability using machine learning techniques. Adv. Intell. Syst. Comput. **1375**(AIST), 412–421 (2021). https://doi.org/10.1007/978-3-030-73050-5_42/COVER
31. Acion, L., Kelmansky, D., Van, L.M.D., Sahker, E., Jones, D.S., Arndt, S.: Use of a machine learning framework to predict substance use disorder treatment success. PLoS ONE **12**, e0175383 (2017). https://doi.org/10.1371/JOURNAL.PONE.0175383
32. Gautam, P., Singh, P.: A machine learning approach to identify socio-economic factors responsible for patients dropping out of substance abuse treatment. Am. J. Publ. Health Res. **8**, 140–6 (2020). https://doi.org/10.12691/ajphr-8-5-2
33. Nath, P., Kilam, S., Swetapadma, A.: A machine learning approach to predict volatile substance abuse for drug risk analysis. In: Proc - 2017 3rd IEEE International Conference on Research in Computational Intelligence and Communication Networks, ICRCICN 2017, pp. 255–258 (2017). https://doi.org/10.1109/ICRCICN.2017.8234516
34. Yetisgen, M., Vanderwende, L.: Automatic identification of substance abuse from social history in clinical text. In: ten Teije, A., Popow, C., Holmes, J., Sacchi, L. (eds.) Artificial Intelligence in Medicine. Lecture Notes in Computer Science(), vol. 10259, pp. 171–181. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59758-4_18/COVER
35. Fede, S.J., Grodin, E.N., Dean, S.F., Diazgranados, N., Momenan, R.: Resting state connectivity best predicts alcohol use severity in moderate to heavy alcohol users. NeuroImage Clin. **22**, 101782 (2019). https://doi.org/10.1016/j.nicl.2019.101782
36. Kumari, D., Swetapadma, A.: Analysis of alcohol abuse using improved artificial intelligence methods. In: Journal of Physics: Conference Series, vol. 1950, p. 012003 (2021). https://doi.org/10.1088/1742-6596/1950/1/012003
37. Oliva, V., et al.: Machine learning prediction of comorbid substance use disorders among people with bipolar disorder. J. Clin. Med. **11**, 1–13 (2022). https://doi.org/10.3390/jcm11143935
38. Surden, H.: Harry surden, artificial intelligence and law: an overview. Georgia State Univ. Law Rev. **35**, 1305–1337 (2019)
39. Macaulay, B.O., Aribisala, B.S., Akande, S.A., Akinnuwesi, B.A., Olabanjo, O.A.: Breast cancer risk prediction in African women using random forest classifier. Cancer Treat. Res. Commun. **28**, 100396 (2021). https://doi.org/10.1016/J.CTARC.2021.100396

40. Haas, O., Maier, A., Rothgang, E.: Machine learning-based HIV risk estimation using incidence rate ratios. Front. Reprod. Health **0**, 96 (2021). https://doi.org/10.3389/FRPH.2021.756405
41. Mbunge, E., et al.: Predicting diarrhoea among children under five years using machine learning techniques, 94–109 (2022). https://doi.org/10.1007/978-3-031-09076-9_9
42. Chingombe, I., et al.: Predicting HIV status using machine learning techniques and bio-behavioural data from the Zimbabwe population-based HIV impact assessment (ZIMPHIA15–16). Cybern. Perspect. Syst., 247–58 (2022). https://doi.org/10.1007/978-3-031-09076-9_24
43. Lee, E.E., et al.: Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. Biol. Psychiatry Cogn. Neurosci. Neuroimaging **6**, 856–864 (2021). https://doi.org/10.1016/J.BPSC.2021.02.001
44. Barenholtz, E., Fitzgerald, N.D., Hahn, W.E.: Machine-learning approaches to substance-abuse research: emerging trends and their implications. Curr. Opin. Psychiatry **33**, 334–342 (2020). https://doi.org/10.1097/YCO.0000000000000611
45. Islam, U.I., Haque, E., Alsalman, D., Islam, M.N., Moni, M.A., Sarker, I.H.: A machine learning model for predicting individual substance abuse with associated risk-factors. Ann. Data Sci., 1–28 (2022). https://doi.org/10.1007/S40745-022-00381-0/METRICS
46. Wray, T.B., Luo, X., Ke, J., Pérez, A.E., Carr, D.J., Monti, P.M.: Using smartphone survey data and machine learning to identify situational and contextual risk factors for HIV risk behavior among men who have sex with men who are not on PrEP. Prev. Sci. **20**, 904–913 (2019). https://doi.org/10.1007/S11121-019-01019-Z/FIGURES/1
47. Bailey, J.D., DeFulio, A.: Predicting substance use treatment failure with transfer learning. Subst. Use Misuse **57**, 1982–1987 (2022). https://doi.org/10.1080/10826084.2022.2125272
48. Carnegie Mellon University. Introduction to Machine Learning. Carnegie Mellon Univ Website (2020)
49. Griffis, J.C., Allendorfer, J.B., Szaflarski, J.P.: Voxel-based Gaussian naïve Bayes classification of ischemic stroke lesions in individual T1-weighted MRI scans. J. Neurosci. Methods **257**, 97–108 (2016). https://doi.org/10.1016/j.jneumeth.2015.09.019
50. Martins, C.: Gaussian naive bayes explained and hands-on with scikit-learn. Tower AI (2022)
51. Tan, P.-N., Steinbach, M., Karpatne, A., Kumar, V.: Introduction to Data Mining. 2nd ed. (2019)
52. Walsh, C.G., et al.: Stigma, biomarkers, and algorithmic bias: recommendations for precision behavioral health with artificial intelligence. JAMIA Open **3**, 9–15 (2020). https://doi.org/10.1093/JAMIAOPEN/OOZ054
53. Marks, C., et al.: Methodological approaches for the prediction of opioid use-related epidemics in the United States: a narrative review and cross-disciplinary call to action. Transl. Res. **234**, 88–113 (2021). https://doi.org/10.1016/J.TRSL.2021.03.018
54. Kyei-Nimakoh, M., Carolan-Olah, M., McCann, T.V.: Access barriers to obstetric care at health facilities in sub-Saharan Africa-a systematic review. Syst. Rev. **6** (2017). https://doi.org/10.1186/s13643-017-0503-x
55. Batani, J., Maharaj, M.S.: Towards data-driven pediatrics in Zimbabwe. In: 2022 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems, pp. 1–7 (2022). https://doi.org/10.1109/ICABCD54961.2022.9855907
56. Hughes, R.A., Heron, J., Sterne, J.A.C., Tilling, K.: Accounting for missing data in statistical analyses: multiple imputation is not always the answer. Int. J. Epidemiol. **48**, 1294–1304 (2019). https://doi.org/10.1093/IJE/DYZ032
57. Gianfrancesco, M.A., Tamang, S., Yazdany, J., Schmajuk, G.: Potential biases in machine learning algorithms using electronic health record data. JAMA Intern. Med. **178**, 1544–1547 (2018). https://doi.org/10.1001/JAMAINTERNMED.2018.3763

58. Wang, Z., et al.: A community-based addiction rehabilitation electronic system to improve treatment outcomes in drug abusers: protocol for a randomized controlled trial. Front. Psychiatry **9** (2018). https://doi.org/10.3389/FPSYT.2018.00556
59. Baingana, F., Al'Absi, M., Becker, A.E., Pringle, B.: Global research challenges and opportunities for mental health and substance-use disorders. Nature **527**, 172–177 (2015). https://doi.org/10.1038/nature16032
60. Winslow, B., Mills, E.: Future of service member monitoring: the intersection of biology, wearables and artificial intelligence. BMJ Mil Health (2023). https://doi.org/10.1136/MILITARY-2022-002306
61. Chitungo, I., Mhango, M., Mbunge, E., Dzobo, M., Musuka, G., Dzinamarira, T.: Utility of telemedicine in sub-Saharan Africa during the COVID-19 pandemic. A rapid review. Hum. Behav. Emerg. Technol. **3**, 843–53 (2021). https://doi.org/10.1002/HBE2.297

View publication stats