

Prediction of Drug Addiction in Pakistan Through Bayesian Modelling

CS 452-L1 PROBABILISTIC GRAPHICAL MODELS

PROJECT PROPOSAL

By

Muminah Khurram	(mk07521)
Syeda Rija Hasan Abidi	(sa07424)
Raza Hashim	(rn07380)

July 13, 2025



DHANANI SCHOOL OF SCIENCE AND ENGINEERING

HABIB UNIVERSITY

FALL 2024

1 Introduction/Problem Statement

Drug abuse and addiction have been identified as a major problem in Pakistan by the World Health Organization (WHO) [1] with 4.25 million drug users identified in 2023 and the number is only growing [2]. With societal neglect, lack of resources for the rehabilitation facilities, and non-existent regulation by authorities, this problem seems to worsen by the year. Generally, the risk of addiction goes unheeded with few tools for practitioners at rehabs such as psychologists and counselors to deduce the likelihood based on the symptoms and demographics of the patient. In this project, we will model the susceptibility of individuals developing drug dependence and addiction, given a multitude of factors such as age bracket, gender, living conditions, psychological health, stress levels, education, etc. We aim to create a Bayesian network with causes and effects pertaining to addiction defined with the help of datasets and domain experts.

2 Relevant Literature Review

Using probabilistic reasoning in the form of Bayesian networks to detect addiction patterns has many different forms in the literature. Gnanasekar et al. use a Bayesian network to create a causal model to predict drug addicts with drug abuse marks and “soft” face biometrics after the classification of images using neural networks. This amalgamation of neural networks with Bayesian networks yielded 84% accuracy in this study. The root node here is “Drug Abuser” and its effect attributes (directed arrow to) are the nodes of Blisters/Acne, Muscle loss, Hair loss, Poor Skin tone, and Gender. The study uses a diagnostic or bottom-up approach where they infer from the images whether the person is an addict or not [3]. Our project complements this study in the opposite direction; Gnanasekar et al. use the symptoms of drug addiction and predict based on those attributes while we aim to predict whether they will fall victim to addiction in the future.

Another study targeting addiction using Bayesian Networks was on Internet Addiction Disorder (IAD) [4]. This study established causal links between IAD and what might be observed as symptoms, we aim to achieve something similar with our study. Doing so would make it easier for stakeholders to identify addiction using all its different manifestations and also identify its severity. Singh et al.’s model achieved satisfactory results, as it made erroneous predictions 5.7% of the time, showcasing the potential of applying Bayesian Networks for a problem very similar to ours. The causal links also aid in establishing a sense of prioritization as to what the most relevant factors are and where the efforts should be focused. The study models 5 causal attributes while one effect attribute — the presence of IAD. Both predictive and diagnostic reasonings are shown with evidence given on the causal and effect attributes. While the primary aim of our study is to predict the susceptibility of individuals to addiction, we can also perform diagnostic reasoning i.e. with the presence of addiction we may deduce the likelihood of their demographics, socio-economic situation, and so on.

A major study that works on the same problem statement as this project is that of Abada et al. where they constructed a Naïve Bayes classification model to predict the likelihood of drug addiction. The parameters considered in the model were demographics, socioeconomic status, proximity to drugs, conflict with law, failure in life, and psychological attributes. This data consists of 26 total attributes and 205 instances. This study yields an accuracy score of 91% indicated by the evaluation metrics of High Detection Rate, False Alarm, Accuracy, Precision, and F-measure [5]. While we will be working with the same attributes as Abada et al., we won’t be considering the attributes to be independent as done in this Naïve Bayes model. However, the evaluation metrics to gauge the

accuracy will remain the same.

We also came across a cross-sectional survey that investigated the rising trend of drug addiction in Pakistan, focusing on sociodemographic factors of patients admitted to rehabilitation centers in Islamabad and Rawalpindi. The study found that most addicts were young, skilled individuals, with a significant portion being educated, with heroin being the most abused drug, and peer pressure along with family disputes serving as key drivers of addiction. The study emphasizes the urgent need for preventive measures and highlights the co-morbid presence of depression in nearly half of the participants as well as challenging the traditional notion that drug addiction primarily affects the uneducated and unemployed, showing that addiction is pervasive across various social strata [6]. This paper ends with a call for better rehabilitation facilities and policymaker intervention to address the alarming rise in substance abuse, we aim to better equip policymakers with relevant contextual data on where preventative measures can be targeted with the use of our model.

[?] For this project, we will integrate and analyze multiple datasets from diverse regions to develop a comprehensive model for understanding and predicting drug addiction trends, with a particular focus on the local context of Pakistan. Our approach leverages a combination of sociodemographic data, addiction risk factors, and behavioral trends to create the Bayesian Network, that fully encapsulates the dependencies and interactions between different variables influencing drug addiction. By incorporating datasets from Pakistan, Bangladesh, the United States, and university student populations, we aim to develop a holistic understanding of the complex dynamics that lead to addiction.

2.1 Dataset Descriptions and Approach

- **Rising Trend of Substance Abuse in Pakistan Data:** This data captures key sociodemographic factors influencing addiction in Pakistan, including age, education, and substance types as taken from patients admitted to rehab. This dataset will serve as a foundation for tailoring our model to local trends, identifying vulnerable groups and high-risk scenarios specific to Pakistan [6]
- **Drug Addiction in Bangladesh Dataset:** This dataset provides valuable information from about 200 individuals, detailing variables such as family relationships, financial status, and emotional/mental health problems. The inclusion of information on conflicts with the law, suicidal thoughts, and exposure to other drug users offers further insight into addiction risk factors. By incorporating these factors as they are coming from a country with similar history, traditions, and socio-economic conditions we can judge their influence on addiction [7].
- **National Survey of Drug Use and Health (NSDUH):** The NSDUH dataset from the U.S. provides a wealth of binary classification variables on drug use and mental health from 2015-2019. With a focus on illicit drugs, alcohol, and tobacco, this dataset will provide a broader global perspective on drug addiction. The dataset's structure will also allow us to test the generalizability of our Bayesian Network across different geographical regions as well as covering enough ground to serve as a baseline for us to build on using the local context [8].
- **Students Drug Addiction Dataset:** The student-focused dataset captures factors related to drug experimentation, academic decline, social isolation, and financial issues. This is particularly relevant as not only does it come from Pakistan but it introduces youth-specific risk factors broadening the scope of our mode [9].

Local Contextualization and Model Design: To develop a Bayesian Network that can address the specific context of Pakistan, we will begin by modeling factors from the NSDUH and Students dataset as our broad baseline capturing basic attributes like age, education, and initial exposure to drugs as well as broader risk factors such as mental health and substance abuse history, to explore how these variables interconnect across different cultural settings. Adding our local context with the Pakistani and Bangladeshi dataset for factors such as family relationships, legal consequences, and co-morbidities like depression, to account for the specific challenges present in South Asian contexts.

3 Domain Expert

Although we have a tremendous amount of data to learn the correlations and probabilities from, we are also in contact with several practitioners and psychologists who will serve as domain experts to help us determine the causal relationships within the variables. They will also determine the probabilities of nodes/variables not covered by the dataset used. All the nodes extracted from the data will be filtered and evaluated in the context of Pakistan with the help of the experts. Since two of the four datasets available were developed in the context of the United States and Bangladesh, many variables of distress, socio-economic factors, and cultural nuances are not accounted for. We aim to fill this gap using the experience and insight of health practitioners at Parwarish Rehabilitation Centre (PRC) and Karwane Hayyat Rehabilitation Center in the areas of Clifton and Kemari of Karachi.

4 Methodology

The first step of our project would be to extract all the relevant variables attached to drug addiction from the data we have collected. To ensure that we have managed to extract all the relevant information and to ensure its correctness we will consult a domain expert to guide us and inform us of any further variables that may be added, this will ensure that the model is as exhaustive as is possible in terms of the variables attached to drug addiction. For establishing the model, after we have identified all the variables we would then have to determine the causal relationships among them. Since we do not have experience of clinical psychology or sociology we will defer to a domain expert to guide us through the process by defining the relationships. Since causal links are the foundation of our project we will need to be extra diligent to not try and do without proper guidance.

The next step then would be to assign meaningful and accurate probabilities to all variables. We believe that we have gathered enough data to allow us to extract probabilities from them. As an added measure to ensure accuracy, we will verify the probabilities we have assigned with the expert in question.

For the creation of the model, we will utilize GeNIe software. Our decision to use GeNIe is based on several factors, including our prior experience with the platform, its user-friendly interface, and its versatility in supporting different variants of Bayesian Networks. Moreover, GeNIe is freely available, making it a cost-effective choice for our purposes.

5 Expected Outcomes

- **Predictive Model for Addiction Susceptibility:** We will develop a Bayesian Network that predicts an individual's susceptibility to drug dependence and addiction based on various factors such as age, gender, education, psychological health, and living conditions to allow healthcare professionals to identify high-risk individuals early, enabling timely interventions.
- **Identification of Key Risk Factors:** Our model will highlight the most influential risk factors contributing to addiction so healthcare professionals and policymakers can focus their efforts on addressing these specific areas.
- **Tool for Mental Health Practitioners:** The Bayesian Network can serve as a decision-support tool for psychologists, counselors, and rehabilitation centers, enabling them to assess a patient's likelihood of developing an addiction to assist in creating individualized treatment plans and prioritizing preventive interventions.
- **Contextualization for Pakistan and Data-Driven Insights for Policymakers:** By focusing on the local data and with the help of our domain expert we will provide insights into how addiction manifests in Pakistan differently from other regions to possibly help local authorities develop contextually relevant preventive measures and in the development of targeted public health campaigns, rehabilitation programs, and policy reforms to combat drug addiction in Pakistan more effectively.

References

- [1] Eastern Mediterranean Health Journal, "Profile of drug users in karachi city, pakistan," *World Health Organization - Regional Office for the Eastern Mediterranean*, 2011. [Online]. Available: <http://www.emro.who.int/emhj-volume-17/issue-1/article6.html>
- [2] United Nations Office on Drugs and Crime (UNODC), "Survey report on drug use in pakistan 2013," 2013. [Online]. Available: https://www.unodc.org/documents/pakistan/Survey_Report_Final_2013.pdf
- [3] S. Gnanasekar and S. Yanushkevich, "Face attributes and detection of drug addicts," in *2019 International Conference on Emerging Security Technologies (EST)*, 2019, pp. 1–6.
- [4] A. Singh and S. Babbar, "Detecting internet addiction disorder using bayesian networks," in *Data Science and Analytics, REDSET 2017*, ser. Communications in Computer and Information Science, B. Panda, S. Sharma, and N. Roy, Eds. Springer, Singapore, 2018, vol. 799.
- [5] W. Abada and A. Bouramoul, "Using machine learning techniques to predict people at-risk for drug addiction: A bayesian-based model," in *2022 4th International Conference on Pattern Analysis and Intelligent Systems (PAIS)*, 2022, pp. 1–7.
- [6] P. Ghazal, "Rising trend of substance abuse in pakistan: A study of sociodemographic profiles of patients admitted to rehabilitation centres," *Public Health*, vol. 167, pp. 34–37, 2019. [Online]. Available: <https://doi.org/10.1016/j.puhe.2018.10.020>

- [7] P. Mostafa, “Drug addiction in bangladesh - reasons,” 2022. [Online]. Available: <https://www.kaggle.com/datasets/protikmostafa/drug-addiction-in-bangladesh-reasons>
- [8] B. Gallamoza, “National survey of drug use and health (2015-2019),” 2015. [Online]. Available: <https://www.kaggle.com/datasets/bgallamoza/national-survey-of-drug-use-and-health-20152019>
- [9] A. Masih, “Students drugs addiction dataset,” 2022. [Online]. Available: <https://www.kaggle.com/datasets/atifmasih/students-drugs-addiction-dataset>