Behavioral and Physiological Insights into Substance Use via Feature Selection and Descriptive Analytics

Zaineh Mihyar
Department of Data Science & Artificial Intelligence, Al Hussein Technical University, Amman, Jordan
22110203@htu.edu.jo

Bassam Kasasbeh
Department of Data Science & Artificial Intelligence, Al Hussein Technical University, Amman, Jordan
Bassam.alkasasbeh@htu.edu.jo

Abstract— Substance use presents complex challenges with profound implications for individuals, families, and public systems. Early detection in high-risk populations is essential for effective intervention. This study employs descriptive analytics to examine behavioral and physiological markers of substance use using data from addiction treatment centers in Amman, Jordan. Behavioral indicators like lying (99.1%), irritability (96.09%), and smoking (93.59%) were prominent, alongside physiological markers such as fatigue pallor (97.65%) and skin redness. Advanced feature selection methods, including SelectKBest and Recursive Feature Elimination (RFE), identified significant predictors of drug use. Correlation and trend analyses revealed key patterns: behavioral traits like reduced work motivation and increased social withdrawal correlated strongly with dependency, while physiological markers reflected the physical toll of addiction. These findings emphasize the value of integrating behavioral and physiological data for evidence-based prevention strategies. This research provides a foundation for targeted interventions and public health frameworks to address the societal burden of substance use.

Keywords— Substance Use Detection, Behavioral Markers, Physiological Indicators, Feature Selection Techniques, Descriptive Analytics

I. INTRODUCTION

Drug use poses a pervasive societal challenge, with profound implications for individuals, families, and communities. It not only undermines physical and mental health but also amplifies socioeconomic disparities and strains public systems, including healthcare and law enforcement [1]. The issue is particularly pronounced among adolescents and young adults, where peer pressure, curiosity, and the propensity for risk-taking converge, often serving as catalysts for initial experimentation with substances. Left unaddressed, such experimentation can escalate into dependency and addiction, further compounding the problem's complexity [2]

Addressing the multifaceted issue of drug use requires a nuanced understanding of its contributing factors and early indicators. Observable behavioral and physiological changes—such as shifts in social interaction, sleeping patterns, and physical health—offer potential markers for identifying individuals at risk [3]. These indicators, when systematically analyzed, can provide crucial insights into the initiation and progression of substance use. Such evidence-based insights enable stakeholders, including families, healthcare professionals, and policymakers, to implement targeted interventions and design prevention strategies.

This study seeks to contribute to the growing body of knowledge on drug use by identifying and analyzing key behavioral and health indicators using descriptive analytics.

Employing advanced techniques such as feature selection and correlation analysis, this research aims to uncover significant patterns and relationships that inform public health strategies. Specifically, the study focuses on:

- **Feature Selection**: Utilizing methods like SelectKBest and Recursive Feature Elimination (RFE) to identify critical variables influencing drug use.

- **Correlation Analysis**: Exploring the interactions between behavioral traits (e.g., lying, decreased motivation) and physiological symptoms (e.g., fatigue, pallor, and skin redness).

- **Descriptive Trends**: Highlighting actionable patterns to enhance understanding and support early detection efforts.

By concentrating on these aspects, the research aims to provide a foundation for proactive measures in mitigating the societal and health-related impacts of drug use. Ultimately, the insights derived from this study will assist in developing evidence-based policies and interventions, fostering resilience within affected communities and promoting long-term well-being.

II. RELATED WORK

Substance use has long been identified as a pressing societal challenge with far-reaching consequences for individuals, families, and public systems. Prior research has consistently highlighted the importance of early identification in mitigating the progression of substance abuse, particularly among vulnerable groups such as adolescents and young adults. Behavioral and physiological markers have emerged as critical indicators for identifying at-risk individuals. For instance, studies have emphasized the interplay between peer pressure, psychological distress, and initial substance experimentation, underscoring the need for comprehensive and multifaceted analytical approaches to address these complexities [3], [4].

Feature selection techniques have gained prominence in behavioral sciences for their ability to isolate relevant variables from complex datasets. Methods such as SelectKBest and Recursive Feature Elimination (RFE) have been widely adopted to enhance the interpretability and performance of predictive models in substance use research. By narrowing down critical indicators, these techniques improve the identification of behavioral precursors to substance use. For example, behaviors such as lying, irritability, and decreased motivation have been identified as significant predictors through these methods [5], [6]. Building on these advancements, the present study applies feature

selection to examine a broader array of behavioral and physiological variables, thereby extending the scope of existing research.

Behavioral markers, including disrupted sleep patterns, diminished motivation, and increased social withdrawal, are consistently associated with early stages of substance use. These markers offer a non-invasive and practical approach for identifying individuals at risk. Prior studies have demonstrated robust links between these behavioral changes and substance dependency, with specific actions such as secrecy and lying emerging as particularly strong indicators [5], [7]. While these findings provide valuable insights into individual behaviors, this study seeks to expand on existing work by exploring interactions between behavioral patterns and physiological changes, offering a more integrated perspective.

In addition to behavioral markers, physiological indicators such as fatigue, pallor, and skin redness have been investigated for their association with substance use. These physical manifestations often accompany behavioral changes, reflecting the physiological burden of drug dependency. For instance, chronic fatigue has been identified as a significant marker in studies exploring the physical effects of long-term drug use . However, much of the existing literature examines these factors in isolation, limiting a holistic understanding of their combined effects. By integrating behavioral and physiological data, this study aims to provide a more comprehensive view of substance use patterns.

Despite the evident value of behavioral and physiological indicators, few studies have explored their combined influence on substance use. Integrative analyses—such as examining correlations between traits like irritability and fatigue—remain underutilized, resulting in fragmented insights. Moreover, many existing frameworks lack the depth required to translate findings into practical prevention strategies. This study addresses these gaps by employing descriptive analytics to uncover actionable patterns, thereby offering a foundation for evidence-based early intervention methods.

While substantial progress has been made in understanding substance use, notable gaps remain in the existing literature. A significant proportion of studies focus heavily on predictive modeling, often overlooking the underlying descriptive trends that inform early detection. Furthermore, the complex interplay between behavioral and physiological factors has not been thoroughly explored. By emphasizing descriptive analytics, this research aims to bridge these gaps, identifying nuanced patterns and relationships that can inform targeted prevention and intervention strategies, particularly for high-risk populations.

III. METHODOLOGY

A. Data Description

Dataset Origin

The dataset used in this study was sourced from addiction treatment centers and clinics in Amman, Jordan [8]. It consists of 184 cases, categorized as follows:

- **155** cases of individuals classified as drug addicts.

- **29** cases of individuals classified as non-addicts.

This dataset provides a robust foundation for analyzing behavioral and physiological factors associated with drug use.

Variables Recorded

The dataset includes **50 variables**, carefully selected to capture a comprehensive range of:

- Behavioral indicators (e.g., irritability, decreased motivation).

- Psychological symptoms (e.g., anxiety, social withdrawal).

- Physiological characteristics (e.g., skin redness, fatigue pallor).

B. Feature Selection

Feature selection was performed using the following methods to identify the most predictive variables:

- **SelectKBest:** This statistical technique evaluates the relevance of individual features by assigning them scores based on their significance in predicting the target variable (drug use). Features with the highest scores were selected.

- **Recursive Feature Elimination (RFE)**: RFE is an iterative method that systematically removes the least significant features while building a predictive model. This process was repeated until an optimized subset of features was identified, maximizing the model's accuracy.

Key Outcomes

Both techniques identified a set of features with consistent predictive significance, including:

- Smoking

- Fatigue pallor

- Irritable mood

- Skin redness

- Decreased motivation to work

These features highlight the interplay of behavioral and physiological indicators strongly associated with drug addiction.

Statistical Analysis for Feature Selection

Statistical validation of feature selection was undertaken by employing SelectKBest and Recursive Feature Elimination (RFE), complemented by Chi-Square tests to ascertain the significance of features related to drug use. The Chi-Square test assessed whether differences in feature distributions between drug users and non-users were statistically significant, thereby confirming the robustness of the feature selection process. Results revealed that several features displayed strong associations with drug use, with p-values significantly below the 0.05 threshold, detailed in Table 1. This evidence supports the hypothesis that behavioral and physiological traits are crucial in identifying substance use. Through the integration of Chi-Square analysis with feature selection techniques, the study ensures that only the most statistically significant and predictive variables are retained for further analysis.

Table 1: Chi-Square Test Results for Feature Association with Drug Use

Feature	Chi-Square (χ^2)	p-value
Smoking	151.90	6.66×10^{-35}
Fatigue Pallor	119.14	9.77×10^{-28}
Increased Personal Expenses	107.68	3.15×10^{-25}
Irritable Mood	106.99	4.46×10^{-25}
Sleeping Habits	105.93	7.63×10^{-25}
Dryness or Redness	100.13	1.42×10^{-23}
Lying	96.52	8.79×10^{-23}
Absence from Work	94.53	2.40×10^{-22}
Dropped Motivation to Work	87.67	7.70×10^{-21}
Decreased Social Interaction	85.61	2.19×10^{-20}

C. Analytical Approach

Descriptive Analytics Techniques

To analyze the dataset, the following descriptive techniques were employed:

- **Correlation Analysis**: Relationships among variables were assessed to identify key factors correlated with drug use. This analysis highlighted individual and combined features contributing to the risk of substance abuse.

- **Behavioral and Physiological Trends**: Prevalence rates for indicators such as "smoking" and "lying" were examined, alongside interaction effects (e.g., "smoking + fatigue pallor") to uncover significant patterns in drug use behaviors.

Focus on Behavioral and Health Correlations

- **Behavioral Indicators**: Variables such as "decreased social interaction" and "irritable mood" were analyzed for their prevalence among individuals classified as drug addicts.

- **Physiological Indicators**: Features like "skin redness" and "fatigue pallor" were examined to identify correlations with substance abuse.

- **Interaction Effects**: Combined feature evaluations (e.g., "smoking" with "dryness or redness") provided deeper insights into the interdependence of behavioral and physiological factors.

D. Visualization and Summary

Descriptive Statistics

Key findings were summarized using percentages and other descriptive metrics, offering a clear understanding of the dataset's characteristics.

Visual Tools

- **Correlation matrix**: Used to visualize the strength of correlations among variables, emphasizing key relationships.

- **Bar Charts**: Illustrated the distribution and prevalence of critical behavioral and physiological indicators within the dataset.

These tools facilitated a comprehensive exploration of the data, contributing to the development of evidence-based intervention strategies targeting drug addiction.

IV. RESULTS

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

A. Feature Selection Outcomes

The application of feature selection techniques, specifically SelectKBest and Recursive Feature Elimination (RFE), identified key variables associated with drug use. SelectKBest highlighted the following features: **smoking**, **dropped motivation to working**, **absence from work**, **sleeping habits**, **decreased social interaction**, **increased personal expenses**, **lying**, **irritable mood**, **dryness or redness**, and **fatigue pallor**. Meanwhile, RFE selected features such as **smoking**, **taken drugs**, **dropped motivation to working**, **commit stole**, **irritable mood**, **weight changed**, **skin redness**, **difficulty completing tasks**, **dryness or redness**, and **fatigue pallor**.

To further validate our feature selection, we conducted **Chi-Square tests**, which confirmed strong statistical associations between several key variables and drug use. The **highest Chi-Square values** were observed for:

- **Smoking (χ^2 = 151.90, p < 0.0001)**

- **Fatigue Pallor (χ^2 = 119.14, p < 0.0001)**

- **Increased Personal Expenses (χ^2 = 107.68, p < 0.0001)**

- **Irritable Mood (χ^2 = 106.99, p < 0.0001)**

- **Sleeping Habits (χ^2 = 105.93, p < 0.0001)**

These findings confirm that behavioral and physiological traits play a crucial role in substance use identification. By integrating **Chi-Square analysis** with feature selection techniques, we ensure that the most **statistically significant** and **predictive** variables are retained for further analysis.

Notably, five features were identified as mutual by both methods, underscoring their critical relevance. These features include **dryness or redness**, **dropped motivation to working**, **smoking**, **fatigue pallor**, and **irritable mood**.

B. Behavioral and Health Insights

Individual Features

1. **Smoking**: Among participants who reported smoking, **93.59%** were identified as drug users. This strong association underscores smoking as a significant behavioral marker for drug use (Figure 1: Smoking vs. Drugs).
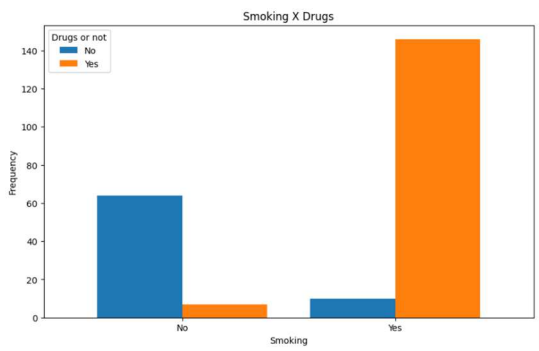
Figure 1: Smoking vs Drugs

2. **Lying**: An overwhelming **99.1%** of participants characterized by habitual lying were also classified as drug users, indicating its potential as a behavioral red flag (Figure 2: Lying vs. Drugs)
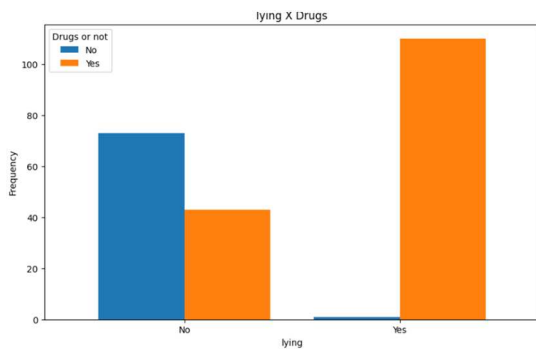

Figure 2: Lying vs Drugs

3. **Irritable Mood**: **96.09%** of individuals exhibiting irritable moods were associated with drug use, further linking psychological states to substance abuse (Figure 3: Irritable Mood vs. Drugs).

Smoking (χ^2 = 151.90, p < 0.0001)

- Smoking is widely recognized as a **gateway behavior** to substance abuse, often preceding the use of illicit drugs.
- Studies suggest that nicotine dependence **alters brain chemistry**, making individuals more susceptible to drug cravings and dependence.

- The strong statistical association in our dataset supports this connection, reinforcing that **smoking is a major risk factor** for drug addiction.

Fatigue Pallor (χ^2 = 119.14, p < 0.0001)

- Fatigue and pallor are **physiological indicators** of substance use, often linked to poor nutrition, sleep deprivation, and chronic drug abuse.
- Many addictive substances (e.g., opioids, stimulants) **disrupt the body's energy metabolism**, leading to persistent fatigue.
- This finding aligns with medical literature, where **fatigue and physical deterioration** are common signs of prolonged drug dependence.

Increased Personal Expenses (χ^2 = 107.68, p < 0.0001)

- Substance addiction frequently leads to **financial instability**, as individuals prioritize drug purchases over basic needs.
- The high Chi-Square value confirms that **economic strain is a strong marker** of substance use, making financial tracking a useful indirect indicator for identifying at-risk individuals.

Irritable Mood (χ^2 = 106.99, p < 0.0001)

- Irritability is a common **psychological symptom** of drug dependence, particularly in withdrawal phases.
- Research shows that **dopamine dysregulation** in addicted individuals leads to mood swings and frustration.
- The statistical significance of this feature reinforces its role as a **behavioral warning sign** of drug addiction.

Sleeping Habits (χ^2 = 105.93, p < 0.0001)

- Drug use **disrupts sleep patterns**, leading to either **insomnia (stimulants like cocaine)** or **excessive drowsiness (opioids, depressants)**.
- Sleep disturbances can also result from **mental health disorders**, which are often co-occurring with addiction.
- Identifying irregular sleep patterns can therefore serve as an **early diagnostic indicator** of substance abuse.
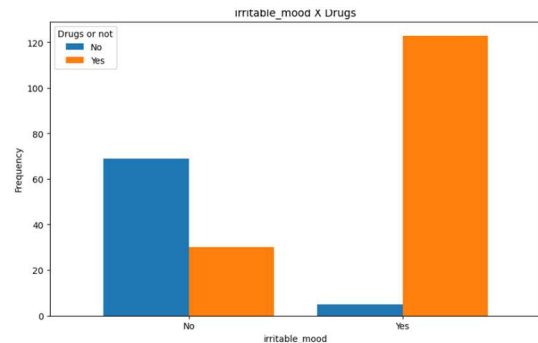

Figure 3: Irritable Mode vs Drugs

Combined Feature Analysis

A deeper exploration of feature interactions revealed notable combined patterns:

- **Smoking, dryness or redness, and drug use**: Nearly **49.8%** of participants who smoked and reported dryness/redness also used drugs (Figure 4: Smoking, Dryness or Redness, and Drug Use).
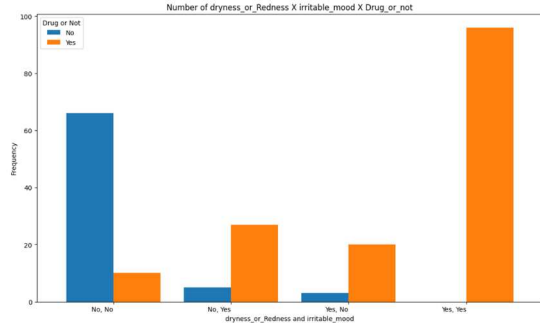

Figure 4: Smoking, dryness or redness, and drug use

- **Smoking and dropped motivation to working**: Among smokers, **50.66%** with dropped motivation to work were drug users (Figure 5: Smoking and Dropped Motivation to Working).
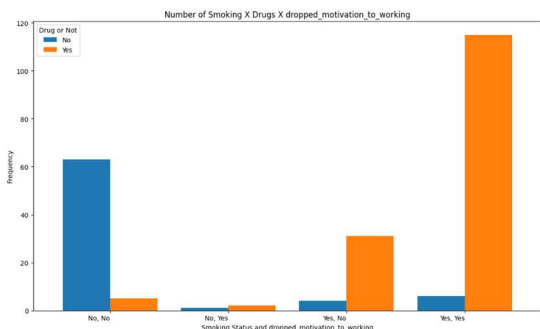

Figure 5: Smoking and dropped motivation to working

- **Fatigue pallor and irritable mood**: For participants exhibiting both fatigue pallor and irritable mood, **45.37%** were drug users (Figure 6: Fatigue Pallor and Irritable Mood).
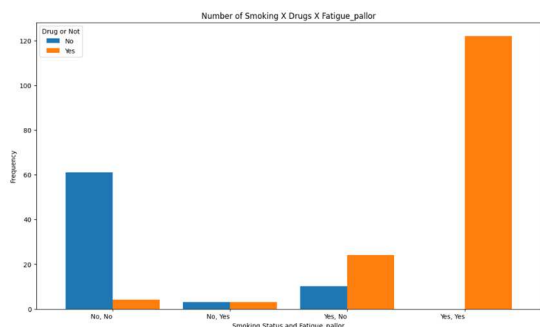

Figure 6: Fatigue pallor and irritable mood

Correlation Analysis

Correlation analysis identified features with strong associations (>70%) to drug use:

- **Smoking**

- **Sleeping habits**

- **Increased personal expenses**

- **Irritable mood**

- **Fatigue pallor**

These correlations suggest that these variables are significant predictors of drug use and warrant further exploration in intervention strategies. Figure 7 presents a correlation matrix highlighting these relationships and their respective strengths.
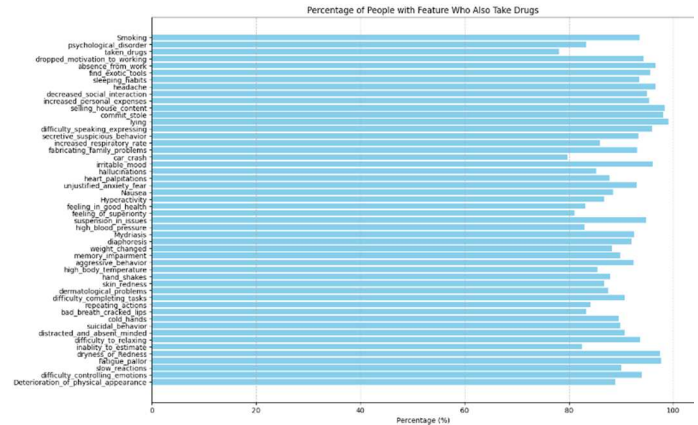

Figure 7: Correlation matrix of features associated with drug use

Patterns and Trends

The analysis revealed several high-association behaviors:

- **Lying** emerged as the most predictive individual behavior, with a near-total association to drug use (Figure 2: Lying vs. Drugs).

- **Fatigue pallor** was observed in **97.65%** of individuals identified as drug users, suggesting its potential utility as a physiological indicator. (Figure 6: Fatigue Pallor and Irritable Mood).

- **Decreased social interaction**: Participants reporting reduced social engagement were associated with drug use in **95%** of cases, emphasizing the social dimension of substance abuse.

These findings provide a nuanced understanding of the behavioral and health factors associated with drug use, highlighting specific markers and patterns that can inform early detection and targeted intervention strategies.

V. DISCUSSION

The analysis of behavioral and physiological factors associated with drug use has yielded critical insights, underscoring the multifaceted nature of substance abuse and its predictors. Among the most notable findings is the robust correlation between smoking and drug use, with 93.59% of smokers in the dataset identified as drug users. This statistic suggests that smoking may function not merely as an isolated behavior but as a potential gateway or concurrent habit associated with drug dependency. Similarly, the high prevalence of fatigue pallor (97.65%) among drug users highlights the physiological toll of substance use, establishing it as a vital marker for early identification and intervention.

This interplay between behavioral traits—such as irritability (96.09%) and deceitful tendencies like lying

(99.1%)—and physiological indicators such as skin redness and fatigue pallor provides a nuanced understanding of substance abuse. These findings extend beyond individual behaviors, suggesting broader health and social consequences of drug dependency. For instance, the observed decrease in work motivation and the increase in personal expenses among users reflect the socioeconomic disruptions that often accompany substance abuse. Together, these markers present a holistic view of the behavioral and physical changes associated with substance use, offering actionable insights for intervention efforts.

From a public health perspective, the findings carry significant implications. Behavioral markers, such as irritability, could inform counseling strategies for individuals exhibiting early signs of substance use. Meanwhile, integrating physiological indicators like fatigue pallor into routine health screenings could enhance early detection efforts. By focusing on at-risk individuals, such as smokers, targeted interventions could mitigate the progression of substance abuse and reduce its broader societal impact.

While the study provides valuable contributions, several limitations must be acknowledged. First, the reliance on self-reported behavioral traits introduces the potential for reporting bias. This limitation is especially relevant given the stigma surrounding substance use, which may influence participants' responses. Although the inclusion of family members for external observations offers additional insights, it also introduces subjective assessments, potentially affecting data reliability.

Second, the dataset's composition—limited to individuals from addiction treatment centers and clinics in Amman, Jordan—restricts the generalizability of the findings. Sociocultural and regional variations in substance use patterns mean these results may not fully represent broader populations. Furthermore, the study's cross-sectional design prevents the establishment of causal relationships between identified markers and drug use. Longitudinal studies would be necessary to determine whether these indicators are precursors to or consequences of substance dependency.

Lastly, while feature selection techniques such as SelectKBest and Recursive Feature Elimination effectively identified significant variables, the analysis primarily focused on descriptive trends. Although these methods provided valuable insights, incorporating advanced predictive modeling could further validate the findings and improve their real-world applicability.

VI. CONCLUSION AND FUTURE WORK

This research identified critical behavioral and physiological markers associated with drug use, including smoking, lying, irritability, fatigue pallor, and skin redness. The combination of SelectKBest and Recursive Feature Elimination (RFE) proved effective in pinpointing these significant variables. Key findings showed that behaviors like lying (99.1%) and irritability (96.09%) were highly predictive of drug use, while physiological markers such as fatigue pallor (97.65%) highlighted the physical toll of substance dependency. These insights contribute to a deeper understanding of the multifaceted nature of substance abuse, emphasizing the value of integrating behavioral and physiological data to support early detection and intervention efforts.

To further enhance the impact of this research, future studies should focus on validating these findings across larger, more diverse populations. Expanding the dataset to encompass different cultural and socioeconomic contexts would improve generalizability. Longitudinal studies could establish causal relationships between the identified markers and the progression of drug dependency. Additionally, integrating advanced predictive models alongside descriptive analytics could refine intervention strategies. Collaborative efforts with healthcare providers to incorporate identified physiological markers into routine screenings and public health initiatives could offer practical tools for early intervention and long-term mitigation of substance abuse impacts.

REFERENCES

[1] L. Manurung, "The Impact of Drug Abuse on Families and Society (Literature Review)," *MSJ Major. Sci. J.*, vol. 2, no. 2, Art. no. 2, May 2024, doi: 10.61942/msj.v2i2.168.

[2] M. Azwar, B. A. Rauf, and A. R. Asrib, "Analysis of Family Behavior, Community Environment, Motivation to Return Drug Users," presented at the 1st World Conference on Social and Humanities Research (W-SHARE 2021), Atlantis Press, Apr. 2022, pp. 210–213. doi: 10.2991/assehr.k.220402.045.

[3] "Promising vulnerability markers of substance use and misuse: A review of human neurobehavioral studies - ScienceDirect." Accessed: Jan. 20, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S002839082100054X?via%3Dihub

[4] S. Zhang, S. Wu, Q. Wu, D. W. Durkin, and F. F. Marsiglia, "Adolescent drug use initiation and transition into other drugs: A retrospective longitudinal examination across race/ethnicity," *Addict. Behav.*, vol. 113, p. 106679, Feb. 2021, doi: 10.1016/j.addbeh.2020.106679.

[5] R. Green *et al.*, "Predictors of Substance Use Initiation by Early Adolescence," *Am. J. Psychiatry*, vol. 181, no. 5, pp. 423–433, May 2024, doi: 10.1176/appi.ajp.20230882.

[6] B. Lees *et al.*, "Promising vulnerability markers of substance use and misuse: A review of human neurobehavioral studies," *Neuropharmacology*, vol. 187, p. 108500, Apr. 2021, doi: 10.1016/j.neuropharm.2021.108500.

[7] R. Green *et al.*, "Predictors of Substance Use Initiation by Early Adolescence," *Am. J. Psychiatry*, vol. 181, no. 5, pp. 423–433, May 2024, doi: 10.1176/appi.ajp.20230882.

[8] M. A. Sukar, A. Sleit, A. I. Abu-Dalhoum, and B. Al-Kasasbeh, "Identifying a Drug Addict Person Using Artificial Neural Networks," *World Acad. Sci. Eng. Technol. Int. J. Comput. Inf. Eng.*, vol. 10, pp. 611–616, 2016.