BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE

PREDICTION OF RELAPSE IN DRUG AND SUBSTANCE ABUSE PATIENTS USING ARTIFICIAL INTELLIGENCE: CASE OF FASAA REHABILITATION CENTER

BY

KUDZAI MAVHUNDUSE

B221229B

SUPERVISOR: MR MHLANGANISO

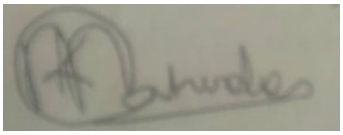
June 2026

Approval Form

The undersigned certify that they have supervised the student **Kudzai Mavhunduse** dissertation entitled **Prediction of Relapse In Drug And Substance Abuse Patients Using Artificial Intelligence: Case Of FASAA Rehabilitation Center** submitted in partial fulfilment of the requirements for the Bachelor of **Science Honours Degree In Software Engineering** at Bindura University of Science Education.

KUDZAI MAVHUNDUSE

STUDENT NAME

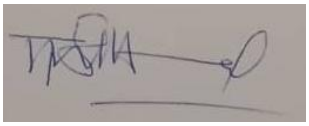


31 MAY 2026

DATE

MR MHLANGANISO

SUPERVISORS

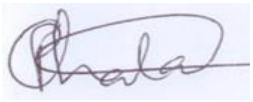


31 MAY 2026

DATE

MR CHAKA

CHAIRPERSON



31 MAY 2026

DATE

EXTERNAL EXAMINER

----------DATE

DEDICATION

I dedicate this project to my beloved parents and family whose love, sacrifices, encouragement, and unwavering belief in me gave me the strength and determination to complete this journey successfully. Your support and prayers continually inspired me to persevere through every challenge encountered during this research.

I also dedicate this work to FASAA Rehabilitation Center for inspiring this study through their commitment to supporting individuals affected by drug and substance abuse. This project was developed with the hope of contributing towards improved patient management, monitoring, and relapse prevention through the use of artificial intelligence and modern technology.

Finally, this work is dedicated to all individuals battling substance abuse, with the hope that technology-driven healthcare solutions may contribute positively toward recovery, rehabilitation, and a better future.

Acknowledgement

First and foremost, I would like to give all glory and gratitude to Almighty God for granting me the strength, wisdom, knowledge, and perseverance throughout the course of this research journey. His guidance and grace enabled me to overcome every challenge encountered during the development of this study.

I would like to express my sincere appreciation to my supervisor, Mr Mhlanganiso, for his continuous guidance, patience, mentorship, and valuable insights throughout this research. Your support and constructive feedback played a significant role in shaping and improving the quality of this project.

My heartfelt appreciation also goes to the lecturers and staff of the Department of Computer Science at Bindura University of Science Education for equipping me with the academic knowledge, practical skills, and professional guidance that made this research possible. Your dedication greatly contributed to my academic and professional growth.

I would also like to extend my sincere gratitude to FASAA Rehabilitation Center for their cooperation, support, and valuable contributions during the development of this system. Their experiences and insights greatly contributed to the understanding of the problem and the design of the proposed solution.

Lastly, I would like to acknowledge everyone who contributed directly or indirectly to the successful completion of this project. Your support and contributions are sincerely appreciated.

ABSTRACT

In this research we concentrated on developing an AI-powered Relapse Prediction System for recovering drug and substance abuse patients to improve continuous patient monitoring, early relapse detection, and timely intervention. To achieve this, we developed a web-based system integrated with machine learning techniques using patient check-in forms, journals, and chatbot interactions to collect and analyze emotional, behavioral, and relapse-related data. Random Forest and Logistic Regression models were trained and tested in Jupyter Notebook to predict emotional distress and relapse risk levels, while automated alerts were sent to social workers whenever high-risk patients were detected. Results show that the chatbot and journal prediction model achieved strong performance with 89% accuracy, precision, recall, and F1-score, demonstrating high effectiveness in detecting emotional distress and behavioral relapse indicators. The check-in form prediction model achieved 61% accuracy, 46% precision, 61% recall, and 52% F1-score, indicating moderate predictive capability in identifying relapse risk levels. Overall, the study demonstrated that integrating artificial intelligence into rehabilitation management can enhance patient support, improve decision-making, and strengthen relapse prevention efforts within rehabilitation centers. The study recommends the adoption of AI-powered rehabilitation systems to improve early relapse detection and continuous patient monitoring in rehabilitation centres. Future improvements should focus on using larger real-world datasets and more advanced AI techniques to enhance prediction accuracy and system reliability.

Table of Contents

Approval Form 2

DEDICATION 3

Acknowledgement 4

ABSTRACT 5

Chapter 1: Problem Identification 9

1.1 Introduction 9

1.2 Background to the Study 9

1.3 Statement of the Problem 11

1.4 Aim of the Study 11

1.5 Research Objectives 11

1.6 Research questions 11

1.7 Research propositions / hypothesis 12

1.8 Justification/significance of the study 12

1.9 Assumptions 13

1.10 Scope/delimitation of the research 13

Chapter 2: Literature Review 13

2.1 Introduction. 13

2.1.1 Definition of Relapse Prediction **in drug and substance abuse treatment** 14

2.2 Relevant theory of the subject matter 14

2.3 Empirical Literature 16

2.4 Research Gap 24

2.5 Chapter Summary 25

Chapter 3: Research Methodology 26

3.1 Introduction 26

3.2 Research Design 26

3.2.1 Data Collection Approaches 27

3.3 Population and Sampling 27

3.3.1 Population 27

3.3.2 Sampling 27

3.3.3 Justification of Sampling Technique 28

3.4 Research Instruments 28

3.4.1 Questionnaires 29

3.4.2 Interviews 29
3.5 Requirements Analysis 29
3.5.1 Functional Requirements 29
3.5.2 Non-Functional Requirements 30
3.5.3 Software Requirements 30
3.5.4 Hardware Requirements 30
3.6 System Development Process 30
3.6.1 System Development Methodology 31
3.6.2 Evolutionary Prototyping 31
3.7 Summary of how the system works 31
3.7.1 Patient Module 31
3.7.2 Social Worker Module 33
3.8 AI Model Training (Jupyter Notebook) 34
3.8.1 Training the Random Forest Relapse Prediction Model used in the Daily Check-in Form 35
3.8.2 Training the Machine Learning model used in Chatbot and Journal 37
3.9 System design 40
3.9.1 Proposed System Flowchart 40
3.10 Testing 42
3.10.1 Black Box Testing 42
3.10.1.1 Testing the Patient Module 42
3.10.1.2 Testing the Social Worker Module 45
3.11 Chapter Summary 47
Chapter 4 - Data Presentation, Analysis and Interpretation 48
4.1 Introduction 48
4.2 Evaluation Measures and Metrix 48
4.2.1 Evaluating the Random Forest Relapse Prediction Model used in the Daily Check-in Form 48
4.2.2 Evaluating the Machine Learning model used in Chatbot and Journal 54
4.3 Analysis and interpretations of results 60
4.3.1 Chatbot and Journal Patient Relapse Prediction Model 60
4.3.2 Check-In Form Relapse Prediction Model 60
4.4 A Summary of Research Findings 61
4.4.1 AI-Powered Relapse Prediction and Early Risk Detection 61
4.4.2 System Usability, User Acceptance, and Reliability 62

4.4.3 Study limitations 62
4.4.3.1 Limited Financial Resources for Data Collection: 63
Chapter Summary 63
CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS 64
5.1 Introduction 64
5.2 Major conclusions drawn 64
5.3 Recommendations 65
5.4 Future Works 65
Summary 65

Chapter 1: Problem Identification

1.1 Introduction

Drug and substance abuse remains a major global public health challenge, placing substantial burdens on individuals, healthcare systems, and societies worldwide due to increasing prevalence rates, rising disability-adjusted life years, and significant socioeconomic costs associated with substance use disorders (Onaolapo et al., 2022). Despite ongoing efforts to improve treatment outcomes, relapse remains a persistent problem in substance abuse treatment, with many individuals returning to substance use after the detoxification process (Nawi et al., 2021). This chapter provides the basis for the current study by giving some background information regarding drug and substance abuse, especially the issue of relapse and relapse prediction. In addition, the chapter explains the statement of the problem and shows what gaps exist in the available substance abuse treatment methods, especially identifying those who can easily relapse.

1.2 Background to the Study

Drug and substance abuse refers to the harmful and continued use of drugs and substances despite their negative social, psychological, physical, or occupational effects (World Health Organization [WHO], 2023). It has become a major global health challenge, affecting individuals, healthcare systems, and social stability worldwide (Onaolapo et al., 2022). The United Nations Office on Drugs and Crime estimates that about 316 million people between 15 and 64 years old were involved in the consumption of drugs in 2023, marking an escalation from the previous years (United Nations Office on Drugs and Crime, 2025). Drug abuse also contributes substantially to morbidity and mortality, with overdose-related deaths continuing to rise globally (Spencer et al., 2022).

In Africa, the availability of prescription and illicit drugs has increased substance abuse, with alcohol, cannabis, and khat being the most commonly abused substances (Odejide, 2006). Cannabis remains the most widely used drug, particularly in West and Central Africa, while the use of amphetamine-type stimulants and benzodiazepines is also increasing (Onaolapo et al., 2022). Africa faces one of the world's largest treatment gaps, as fewer than one in ten individuals requiring treatment receive adequate care (UNODC, 2025).

Drug and substance abuse is becoming rampant in Zimbabwe, to such an extent that there is need for concern about its effects on the security and health of the nation (Government of Zimbabwe & UNICEF Zimbabwe, 2024). The MoHCC (2024) stated that the illicit drugs and narcotics have become one of the challenges facing the country's younger generation. According to the research conducted by Matutu and

Mususa (2019), over half (66%) of youths between 10 and 24 years of age from urban high-density communities used illicit drugs within the past three months before the study. According to studies conducted by ZCLDN, NAC, and MoHCC (2019), injectable drugs are increasingly popular among youths; crystal meth, and pharmaceutical drugs have been identified as popular drugs. This trend has resulted in the increase in the number of mental health admissions where drug-induced psychosis constituted over 80% of the patients admitted.

Despite efforts to improve prevention and treatment programs, relapse remains a major challenge in substance abuse recovery (Stillman & Sutcliff, 2020). Relapse is influenced by biological, psychological, and social factors such as cravings, stress, poor social support, and limited access to continuous care (Witkiewitz & Marlatt, 2004). In Africa, relapse rates are worsened by understaffed treatment centers, poor follow-up systems, and lack of integrated health information systems (Odejide, 2006; UNODC, 2025). Current relapse prevention methods mainly depend on counseling sessions and self-reporting, yet patients may hide important information because of stigma or fear of judgment, reducing the effectiveness of early intervention (Tai et al., 2012).

Studies by Ghitza and Spenenberg (2011) show that relapse is often linked to systemic and technological limitations within treatment facilities rather than individual failure alone. Many treatment systems are reactive, responding only after relapse occurs instead of identifying warning signs early. Zimbabwe faces similar challenges due to fragmented treatment services and lack of structured patient monitoring systems. Frimpong et al. (2023) further note that fragmented digital tools limit continuous patient monitoring and timely intervention.

Recent studies suggest that integrating digital health technologies and Artificial Intelligence (AI) into substance abuse treatment can improve relapse prevention (Marsch, 2020). AI-powered systems can analyze patient data, identify relapse patterns, and predict high-risk behaviors earlier than traditional methods, enabling healthcare providers to make timely and informed decisions.

Given the growing burden of drug and substance abuse and the high risk of relapse in Zimbabwe, there is a need for innovative and data-driven treatment approaches. This study aims to develop a drug and substance abuse patient relapse prediction system for FASAA Rehabilitation Centre. The proposed system will use machine learning techniques to analyze patient-reported data such as mood, cravings, and personal reflections to identify relapse risk patterns. When high-risk patterns are detected, the system will automatically alert caregivers to support timely intervention and improve patient recovery outcomes

1.3 Statement of the Problem

In Zimbabwe, drug and substance abuse has placed significant pressure on the mental health system, with approximately 80% of psychiatric hospital admissions linked to drug-induced conditions, resulting in high treatment costs and repeated cycles of care (MoHCC, 2024). Determining an individual's risk of relapse remains a major challenge, as treatment decisions largely rely on counseling sessions in which many patients withhold critical information due to fear, stigma, or mistrust. Consequently, healthcare providers are often unable to accurately identify patients at high risk of relapse in time, leading to preventable relapses, repeated hospitalizations, and continued strain on already limited drug and substance abuse treatment services.

1.4 Aim of the Study

This study is carried out to bring a solution that ensures that patients have fully recovered before they get discharged from FASAA rehab. It has been noted that there are various underlying factors that causes relapse after the detoxification process among Drug and Substance Abuse (DSA) patients. These include psychological and psycho social factors that patients do not discuss with their care givers due to fear of stigma among other reasons. This study aims to bring a solution by designing a machine learning relapse prediction module. The AI module will interact with patients on daily basis, patients will log in their states, craving levels, personal reflections and the AI with provide suggestive solutions to make them feel better, at the same time it will send an alert to their caregivers if a patient indicates high risk of relapse.

1.5 Research Objectives

i. To research on the techniques that can be used to predict relapse in drug and substance abuse patients
ii. To develop a Random Forest and NLP-based machine learning algorithm to predict relapse of DSA patients at FASAA.
iii. To determine how efficient the random forest model and NLP model can be in predicting relapse of DSA patients at FASAA

1.6 Research questions

i. What techniques can be used to predict relapse in DSA patients?
ii. How can a Random Forest model and NLP model be developed to predict relapse in DSA patients at FASAA?

iii. How effective is the Random Forest and NLP model in predicting relapse in DSA patients at FASAA?

1.7 Research propositions / hypothesis

H1a: Rehabs implementing an AI powered relapse prediction system will more accurately identify patients at high risk of relapse than centers relying on manual methods

H1b: Patients receiving care at the rehabilitation center using the AI-powered relapse prediction system will have lower relapse rates than patients receiving care under standard treatment protocols.

1.8 Justification/significance of the study

1.8.1 To the Student

This study will enhance the student's understanding of DSA while strengthening practical skills in software development and artificial intelligence. It provides an opportunity to apply theoretical knowledge to the development of a real-world, technology-driven solution addressing a critical national health challenge of DSA in Zimbabwe.

1.8.2 To the University board of knowledge

This study will strengthen the university's research portfolio by providing a rigorous case study on the application of artificial intelligence in drug and substance abuse treatment setting. It will generate valuable empirical insights and foundational data to support future scholarly work in drug and substance abuse within Zimbabwe and Southern African context.

1.8.3 To Rehabilitation centers in Zimbabwe

This study addresses a critical operational gap by providing an AI-powered relapse prediction system. By generating data-driven, early-warning alerts, this solution will equip FASAA with a tool to shift from reactive crisis management to proactive, personalized intervention. This capability is vital to minimizing relapse rates and improving long-term recovery outcomes in a system where up to 80% of admissions are already linked to drug-induced conditions (MoHCC, 2024).

1.9 Assumptions

The study is based on the following assumptions that:

i. Sufficient, relevant, reliable and real-time DSA patient data (e.g., mood logs, craving intensity, treatment attendance, demographic factors) can be collected from FASAA rehabilitation center to effectively train and validate the relapse prediction machine learning model.
ii. Measurable DSA patient-reported data and behavioral indicators (e.g., increased craving frequency, negative mood patterns, social connectivity changes) have a consistent and identifiable relationship with subsequent relapse events.
iii. The actionable alerts generated by DSA relapse predictive model will enable more effective, timely interventions
iv. Shifting clinical practice at FASAA from being reactive to being proactive, will contribute to a measurable reduction in relapse rates.

1.10 Scope/delimitation of the research

Delimitations refer to the intentional boundaries and choices made by the researcher to define the scope of the study (Theofanidis & Fountouki, 2019). In this study, the researcher intentionally defined the scope by choosing to focus only on aspects related to relapse prediction. The research focuses on FASAA Rehabilitation Center in Zimbabwe and is limited to the design and development of an AI-powered relapse prediction system. The system is confined to three core modules, namely the staff module, patient module, and administrator module, which are designed to support patient treatment trough relapse prediction. Although many operational aspects exist within the rehabilitation center, this study will focus only on the staff involved in therapeutic and care activities, patients, and administrative modules relevant to patient management and relapse prevention. It will exclude other operational units of the rehabilitation center, such as kitchen services, pharmacy operations, accounting, or advanced financial management systems, which are outside the scope of this research. In addition, the study is limited to data and activities observed over a 4-month period, from January 2025 to April 2026, to provide a clear temporal scope for the research.

Chapter 2: Literature Review

2.1 Introduction.

This chapter examines relevant theories, empirical studies, and scholarly literature related to relapse prediction in drug and substance abuse treatment. Understanding the factors associated with relapse, as well

as the methods used to predict it, is essential for the development of an effective AI-powered relapse prediction module for drug and substance abuse patient treatment (Vafaie and Kober, 2022). Through a critical analysis of existing research, this review seeks to identify best practices, expose limitations in current relapse prediction approaches, and establish a conceptual and empirical foundation for system development

2.1.1 Definition of Relapse Prediction **in drug and substance abuse treatment**

Relapse prediction refers to the process of estimating the likelihood that an individual will return to previous drug or substance use after a period of abstinence, despite an intention to remain abstinent (Efthimiou et al., 2024). The process involves the use of multiple techniques to assess an individual's risk of returning to drug or substance use during the recovery process. These techniques may include the collection and analysis of clinical history, behavioral patterns, psychological assessments, and substance use records, as well as the application of statistical or computational models to identify individuals at higher risk of relapse.

2.2 Relevant theory of the subject matter

2.2.1 Cognitive Behavioral Theory

Cognitive Behavioural Theory (CBT) provides an important framework for understanding human behaviour as the result of interactions between thoughts, emotions, and actions (Beck, 2020). The theory explains that negative or distorted thinking patterns play a major role in shaping harmful behaviors, including relapse in drug and substance abuse treatment (Marlatt & Donovan, 2005). Because of this, CBT has been widely used in addiction treatment to help individuals recognize high-risk situations, develop effective coping strategies, and change unhealthy behavioral patterns (Carroll & Kiluk, 2017). Research shows that CBT remains one of the most effective approaches in substance abuse treatment, particularly in identifying triggers and cognitive distortions linked to relapse (McHugh, Hearon, & Otto, 2010). Importantly, CBT views relapse not as a sudden event, but as a gradual and predictable process influenced by psychological and environmental factors such as stress, emotions, and social interactions (Witkiewitz & Marlatt, 2004).

Recent empirical evidence further supports the effectiveness of CBT in relapse prevention. A study conducted by Rakhshanderou et al. (2025) applied a cognitive-behavioural intervention based on relapse prevention principles to individuals who smoke. The findings showed significant reductions in lapse and relapse rates among participants who received the intervention compared to those who did not. In addition,

participants demonstrated improved coping skills, increased confidence in resisting substance use, and reduced exposure to high-risk situations (Rakhshanderou et al., 2025). These results highlight the practical value of CBT in managing and preventing relapse.

In this study, CBT provides a strong theoretical foundation for understanding how patients' thoughts and behaviors influence relapse risk. The proposed system applies this theory by collecting relevant behavioral and psychological data, such as mood changes, treatment adherence, and exposure to triggers. This information is then analyzed by the AI model to predict the likelihood of relapse. Therefore, CBT not only guides the type of data collected but also supports how the system interprets patient behaviour, ensuring that the technological solution is grounded in established psychological principles.

2.2.2 Relapse Prevention Theory

Relapse Prevention (RP) Theory was originally developed by Marlatt and Gordon in 1985 to provide a comprehensive cognitive-behavioral framework for understanding and managing the recurrence of substance use following treatment. The theory conceptualizes relapse not as a sudden failure of willpower but as a dynamic and multi factorial process involving the interaction of interpersonal and interpersonal determinants. According to Marlatt & Donovan, 2005 central to this model is the identification of high-risk situations, which include emotional distress, environmental triggers, and social pressures that reduce an individual's sense of control and increase vulnerability to substance use. The theory further explains that an individual's response to these high-risk situations determines the likelihood of relapse, where effective coping strategies enhance self-efficacy and reduce risk, while ineffective coping may lead to a lapse (Marlatt & Donovan, 2005). This initial lapse can trigger the Abstinence Violation Effect, a psychological response characterized by guilt and loss of control, which may escalate into full relapse if not properly managed (Larimer et al., 1999). Contemporary perspectives extend the theory by emphasizing that relapse is influenced by a combination of factors such as craving, motivation, mental health conditions, and past substance use history, all of which interact in a complex and dynamic manner (Witkiewitz & Marlatt, 2004). As a result, the theory supports the use of continuous monitoring and data-driven approaches in identifying relapse risk, making it highly relevant to this study, which seeks to apply artificial intelligence techniques in predicting and preventing relapse among patients undergoing substance abuse treatment.

2.2.3 Predictive Analytics and Machine Learning Theory

Predictive Analytics and Machine Learning (ML) Theory focuses on the use of historical data, statistical techniques, and algorithms to identify patterns and predict future outcomes (Simon & Aliferis, 2024). As a branch of artificial intelligence, machine learning enables computers to learn from data and make informed

decisions without being explicitly programmed. These techniques are particularly effective in analyzing complex datasets, such as electronic medical records, where they can uncover patterns that may not be easily detected by human clinicians (Seringa et al., 2024).

In healthcare, machine learning is increasingly applied in areas such as risk prediction, early diagnosis, treatment optimization, and clinical decision support (Santoso et al., 2025; Yang et al., 2025). Recent developments show that AI-driven prediction models have evolved from simple, single-condition analyses to more advanced systems capable of integrating multiple data sources, including demographic, clinical, and behavioural data, into comprehensive predictive frameworks (Yang et al., 2025). This advancement has made predictive analytics a critical component of modern healthcare systems. Empirical studies further demonstrate the effectiveness of machine learning in predicting patient outcomes. For example, Xu et al. (2025) conducted a large-scale study involving 5,255 individuals with substance use disorders and found that the XGBoost model achieved strong predictive performance, with an area under the curve of 0.76. The study identified key predictors of relapse, including duration of substance use, type of substance, dosage, and age of first use (Xu et al., 2025). These findings highlight the ability of machine learning models to identify important risk factors and support early intervention.

In this study, Predictive Analytics and Machine Learning Theory provides the core technical foundation for the AI-powered relapse prediction system. The system applies machine learning algorithms to analyse patient data and generate relapse risk predictions, enabling healthcare providers to make proactive and informed decisions. By leveraging these techniques, the proposed system enhances early detection of relapse risk and supports personalized treatment planning, thereby improving patient outcomes.

2.3 Empirical Literature

Recent advances in artificial intelligence and data-driven approaches have significantly improved the prediction of relapse in substance use disorders (SUD), particularly through the use of real-time and behavioral data. A study by Wyant et al. (2024) explored the application of machine learning techniques using Ecological Momentary Assessment (EMA) data to predict alcohol relapse. The researchers developed predictive models capable of forecasting relapse across multiple temporal windows, including weekly, daily, and hourly predictions. The study highlights that relapse in individuals with Alcohol Use Disorder is a dynamic and multifactorial process, influenced by continuously changing psychological and contextual factors such as craving, stress, and self-efficacy (Wyant et al., 2024). Traditional approaches that rely solely on past behavior were found to achieve only moderate predictive performance, demonstrating limitations in capturing the complexity of relapse (Wyant et al., 2024). The researchers adopted a data-driven

quantitative approach, utilizing EMA data collected multiple times per day from participants in early recovery to train and validate predictive models. The findings reveal that advanced models incorporating multiple behavioral and contextual features significantly outperform baseline models, achieving excellent predictive performance with high accuracy across all time windows (Wyant et al., 2024). Notably, the study found that shorter prediction windows, particularly hourly predictions, yielded superior performance due to their ability to capture immediate fluctuations in relapse risk. Furthermore, the research demonstrates that relapse risk is highly individualized, with different factors influencing different individuals at specific moments in time, thereby emphasizing the importance of personalized interventions (Wyant et al., 2024). The study proposes the integration of these predictive models into smart digital therapeutic systems capable of delivering just-in-time, adaptive interventions tailored to a patient's current risk profile (Wyant et al., 2024). However, the research also identifies several challenges, including limited demographic diversity, the burden of frequent data collection, and the trade-off between prediction sensitivity and precision, particularly in real-time predictions. The authors recommend further rsesearch to improve model generalizability, reduce user burden through passive data collection, and enhance the effectiveness of intervention strategies in real-world settings.

Similarly, Curtis et al. (2023) introduced artificial intelligence-based approach for predicting substance use disorder (SUD) treatment outcomes using social media data. The study was motivated by persistent challenges in SUD treatment, particularly high rates of relapse and treatment dropout, as well as limitations associated with traditional intake assessments that rely on retrospective self-report data (Curtis et al., 2023). To address these challenges, the researchers developed a digital phenotyping framework using patients' historical social media language, applying a deep learning model, Bidirectional Encoder Representations from Transformers (BERT), to extract meaningful behavioral and psychological patterns (Curtis et al., 2023). This approach enabled the creation of language-based digital phenotypes that reflect real-time, ecologically valid aspects of patients' lived experiences prior to treatment intake (Curtis et al., 2023). The system demonstrated superior predictive performance compared to standard psychometric tools such as the Addiction Severity Index (ASI), achieving higher accuracy in predicting 90-day treatment outcomes, including relapse and dropout (AUC = 0.725 versus 0.658). Furthermore, the combined use of digital phenotypes and ASI data improved predictive performance even further (AUC = 0.739), highlighting the complementary value of AI-driven and traditional assessment methods (Curtis et al., 2023). This system enhances early risk identification by enabling clinicians to detect high-risk individuals at the point of intake, thereby supporting targeted interventions and improved treatment planning (Curtis et al., 2023). However, despite its effectiveness, the approach is limited by its reliance on static, pre-treatment social media data, which may not fully capture the dynamic and evolving nature of relapse risk during treatment. Additionally,

the model depends on the availability of sufficient social media data and may not generalize across different populations or cultural contexts.

While previous studies focused on AI-driven and real-time behavioral data, Lauvsnes, Gråwe and Langaas (2022) examined relapse prediction using intensive longitudinal data on mental health, cognition, and craving. Substance use disorders are characterized by high relapse rates, ranging between 40–60%, with a significant proportion of relapses occurring within the first year after treatment (Lauvsnes, Gråwe and Langaas, 2022). Although prior research has identified key predictors such as craving, impaired self-control, and mental health distress, there remains limited understanding of how these factors interact dynamically over time to influence relapse (Lauvsnes, Gråwe and Langaas, 2022). The study addresses this gap by investigating the temporal and prospective relationships between mental health symptoms, executive self-control, craving, and relapse among individuals undergoing inpatient treatment for polysubstance use disorders (Lauvsnes, Gråwe and Langaas, 2022). The researchers employed an intensive longitudinal design using repeated momentary assessments to capture within-subject variability in mental health, cognition, and craving, allowing for a more accurate understanding of real-time fluctuations (Lauvsnes, Gråwe and Langaas, 2022). The findings revealed that mental health symptom intensity predicts subsequent self-control, variability in self-control predicts future craving intensity, and variability in craving is a strong predictor of subsequent relapse (Lauvsnes, Gråwe and Langaas, 2022). Notably, an increase in craving variability was associated with a 100% increase in relapse risk (Lauvsnes, Gråwe and Langaas, 2022). Furthermore, the study demonstrated that within-subject variability is a more sensitive and reliable predictor of relapse compared to static group-level measures (Lauvsnes, Gråwe and Langaas, 2022). These results highlight that relapse is driven by dynamic, temporally linked changes in psychological and cognitive states rather than stable traits alone. The study contributes significantly to the field by showing that continuous monitoring of mental health distress, cognitive functioning, and craving can improve relapse prediction and enable personalized, just-in-time interventions (Lauvsnes, Gråwe and Langaas, 2022). Despite providing valuable insights into the temporal predictors of relapse, the study is limited by lack of a real-time, deployable system, highlighting the need for AI-driven solutions that enable continuous monitoring and just-in-time intervention in substance use disorder management.

Similarly, Farooq and Riaz (2022) utilized artificial neural networks to examine psychosocial predictors of relapse, identifying risk-taking behavior, social conformity, and emotional instability as key determinants. The study highlights that relapse remains a significant global challenge, with more than 60% of individuals returning to substance use after treatment, largely due to a combination of psychological and social influences (Farooq and Riaz, 2022). These factors include low self-esteem, anxiety, depression, poor decision-making, risk-taking behavior, and social pressures such as peer influence and lack of family

support (Farooq and Riaz, 2022). The primary aim of the study was to identify and classify the psychosocial factors that predict relapse among individuals undergoing rehabilitation. The researchers adopted a cross-sectional research design, collecting data from 200 participants in rehabilitation centers using a standardized Psychosocial Functioning Scale, and applied artificial neural network analysis to examine the predictive relationships between these factors and relapse (Farooq and Riaz, 2022). The findings revealed that multiple psychosocial variables significantly contribute to relapse, with risk-taking behavior emerging as the most influential predictor, followed by social conformity, decision-making confidence, and childhood-related problems. The study further established that relapse is influenced by emotional instability, negative mood states, poor social support, and exposure to high-risk environments, emphasizing the complex and individualized nature of relapse behavior (Farooq and Riaz, 2022). Additionally, the results indicate that individuals who engage in multiple substance use and those exposed to socially permissive environments are at a higher risk of relapse (Farooq and Riaz, 2022). However, the study also identifies several limitations, including a restricted sample size and limited geographic coverage, which affect the generalizability of the findings. The authors recommend further research involving more diverse populations to better understand relapse dynamics and to support the development of effective intervention strategies.

In contrast, Yamashita, Yoshioka and Yajima (2021) focused on resilience as a predictor of relapse using statistical techniques such as correlation and regression analysis. The study utilized standardized instruments, including the Stimulant Relapse Risk Scale (SRRS) and the Bidimensional Resilience Scale, alongside statistical techniques such as correlation and multiple regression analysis to examine associations between patient characteristics and relapse risk (Yamashita, Yoshioka and Yajima, 2021). The findings revealed a significant negative relationship between resilience and relapse risk, with higher levels of acquired resilience associated with lower likelihood of relapse ($r = -0.314$, $p < 0.01$) (Yamashita, Yoshioka and Yajima, 2021). Additionally, employment status emerged as a significant predictor, with employed individuals demonstrating reduced relapse risk (Std. $\beta = -0.446$, $p < 0.05$) (Yamashita, Yoshioka and Yajima, 2021). These results highlight the importance of psychosocial stability and adaptive coping mechanisms in supporting recovery, emphasizing that resilience is not a fixed trait but one that can be developed through environmental and therapeutic interventions (Yamashita, Yoshioka and Yajima, 2021). The study further underscores the role of continuous treatment, social support systems, and structured recovery programs in enhancing resilience and reducing vulnerability to relapse (Yamashita, Yoshioka and Yajima, 2021). However, despite its valuable contributions, the study is limited by its small sample size and cross-sectional design, which restrict its ability to capture the dynamic and evolving nature of relapse risk over time. Furthermore the absence of advanced predictive techniques limits its applicability in real-time clinical decision-making.

Similarly, Faltushanskiy et al. (2023) investigated personality traits using the NEO-PI-R framework and found that traits such as Conscientiousness reduce relapse risk, while Impulsiveness increases vulnerability. The study employed an observational research design using a sample of 441 participants recruited from a specialized treatment center for professionals (Faltushanskiy et al., 2023). Personality traits were assessed at the beginning of treatment using the NEO-PI-R, which measures the Big Five personality domains and their individual facets, while relapse was monitored over a one-year period post-treatment (Faltushanskiy et al., 2023). The researchers applied logistic regression analysis, including both univariate and multivariate models, to examine the relationship between personality traits and relapse while controlling for demographic variables. The findings revealed that 39% of participants relapsed within one year, confirming the persistent challenge of relapse in SUD treatment. At the domain level, Neuroticism was associated with a higher likelihood of relapse, while Agreeableness and Conscientiousness were associated with reduced relapse risk; however, after controlling for demographic factors, only Conscientiousness remained a significant predictor (Faltushanskiy et al., 2023). More importantly, the study provided deeper insights at the facet level, identifying Impulsiveness as a significant predictor of increased relapse risk, while facets such as Dutifulness, Competence, Self-Discipline, and Straightforwardness were associated with lower relapse rates. These findings highlight the importance of examining personality at a more granular level rather than relying solely on broad trait categories. Despite its contributions, the study has several limitations, including a lack of diversity in the sample, reliance on a single treatment site, and the use of traditional statistical techniques that may not fully capture the complex, non-linear interactions between multiple relapse predictors. Furthermore, the predictive approach used is static and not integrated into real-time clinical decision-making systems. This creates a significant research gap, as there is a need for more advanced, data-driven approaches capable of handling complex and dynamic patient data.

Moreover, Yuan et al. (2023) applied computational modeling techniques to assess decision-making processes in predicting relapse among individuals with alcohol use disorder. The researchers employed a comparative and longitudinal design involving 50 clinically diagnosed AUD patients and 46 healthy controls, utilizing the Balloon Analogue Risk Task (BART) to assess risk-taking behavior (Yuan et al., 2023). In addition to traditional behavioral measures such as adjusted average pumps, the study integrated advanced computational modeling techniques, specifically the exponential-weight mean–variance (EWMV) model and a four-parameter BART model, to capture underlying cognitive processes influencing decision-making. Psychological and clinical variables, including anxiety, depression, impulsivity, and alcohol craving, were also collected, and participants were followed up for three months to determine relapse outcomes (Yuan et al., 2023). The findings revealed that while traditional behavioral measures failed to show significant differences between groups, the computational model identified significantly higher risk preference among AUD patients, particularly those who relapsed. Furthermore, risk preference

emerged as a significant predictor of relapse, achieving a moderate predictive performance (ROC-AUC = 0.72), thereby demonstrating the superiority of computational approaches over conventional methods in capturing subtle cognitive impairments. The study concluded that individuals with higher risk-taking propensity are more vulnerable to relapse, highlighting the importance of computational biomarkers in addiction research. However, despite its contributions, the study is limited by a relatively small and gender-restricted sample, short follow-up duration, and reliance on a single behavioral task, which may affect generalizability and predictive robustness. Additionally, the absence of real-time data integration and lack of implementation within an intelligent clinical decision-support system indicate a significant research gap. Therefore, this creates an opportunity for the development of more robust, AI-powered relapse prediction systems that integrate diverse patient data sources and provide continuous, personalized monitoring and intervention, as proposed in the current study.

Furthermore, Kabisa et al. (2021) investigated relapse prevalence and associated factors using retrospective clinical data. Using a sample of 391 hospitalized patients, the researchers found a high relapse prevalence of 59.9%, with most cases occurring within a short period after discharge, particularly between two weeks and three months. The study identifies multiple contributing factors, including socio-demographic, environmental, interpersonal, and psychological variables, and employs statistical techniques such as bivariate and multivariate logistic regression to determine their significance. The findings reveal that relapse is strongly associated with peer group influence, family conflict, psychological stress, accessibility of substances, and the use of multiple substances, while variables such as age, gender, and education level were not statistically significant (Kabisa et al., 2021). The results further demonstrate that patients with shorter hospitalization periods, those living in environments with strong peer or dealer influence, and individuals lacking adequate parental support particularly those not living with both parents are at a higher risk of relapse. The study addresses critical gaps in relapse prevention by emphasizing that treatment approaches focusing solely on detoxification are insufficient, as relapse is influenced by a complex interaction of behavioral, social, and environmental factors. It highlights the importance of extended rehabilitation, continuous follow-up, and supportive environments in improving recovery outcomes. Additionally, the research underscores the role of family involvement and social support systems in reducing relapse, while also pointing out that exposure to drug-using environments significantly undermines recovery efforts. Despite its contributions, the study acknowledges limitations such as its focus on a single rehabilitation center and reliance on retrospective data, which may affect generalizability and limit causal inferences. Nevertheless, the findings provide valuable insights that can inform the development of comprehensive relapse prevention strategies, including community-based interventions, policy improvements, and enhanced post-treatment support systems. These insights are particularly relevant

for informing future research and designing effective interventions aimed at reducing relapse rates and improving long-term recovery among individuals with substance use disorders.

Year	Author	Technique Applied	Results
2024	Wyant et al.	Machine Learning using Ecological Momentary Assessment (EMA) data (real-time behavioral and contextual data; multi-time window prediction: hourly, daily, weekly)	Achieved high predictive accuracy for relapse, with shorter (hourly) prediction windows performing best. Advanced models outperformed traditional approaches, showing relapse is dynamic and individualized. Highlighted importance of personalized, real-time interventions but noted challenges such as limited diversity and data collection burden.
2023	Curtis et al.	Deep Learning (BERT) applied to social media data for digital phenotyping; integration with traditional assessment (Addiction Severity Index - ASI)	Achieved higher predictive accuracy for 90-day treatment outcomes (AUC = 0.725) compared to ASI alone (AUC = 0.658). Combined model improved performance further (AUC = 0.739). Enabled early risk detection but limited by reliance on static pre-treatment social media data and generalizability concerns.
2022	Lauvsnes, Gråwe & Langaas	Intensive Longitudinal Data Analysis using repeated momentary assessments (mental	Found that dynamic changes in mental health, self-control, and craving strongly predict relapse. Craving variability increased relapse risk by 100%. Within-

		health, cognition, craving dynamics)	subject variability proved more accurate than static measures. Highlighted importance of continuous monitoring but lacked a real-time AI-based implementation system.
2022	Farooq & Riaz	Artificial Neural Networks (ANN) applied to psychosocial data (cross-sectional study using Psychosocial Functioning Scale)	Identified key predictors of relapse including risk-taking behavior (most significant), social conformity, emotional instability, and poor decision-making. Showed relapse is influenced by psychological and social factors, with over 60% relapse rate. However, limited by small sample size and lack of generalizability.
2021	Yamashita, Yoshioka & Yajima	Statistical analysis (Correlation and Multiple Regression) using standardized scales (SRRS and Bidimensional Resilience Scale)	Found a significant negative relationship between resilience and relapse risk ($r = -0.314$). Higher resilience and employment status reduced relapse likelihood. Highlighted importance of psychosocial stability, but limited by cross-sectional design and lack of real-time predictive capability.
2023	Faltushanskiy et al.	Logistic Regression (univariate and multivariate) using NEO-PI-R personality assessment framework	Identified personality traits influencing relapse: Conscientiousness reduced risk while Impulsiveness increased it. 39% relapse rate observed within one year. Facet-level traits (e.g., self-discipline) provided deeper

			insights. Limited by lack of diversity, single-site data, and static predictive approach.
2023	Yuan et al.	Computational Modeling (EWMV model and four-parameter BART model) combined with behavioral task (BART) and longitudinal analysis	Computational models outperformed traditional behavioral measures, identifying higher risk preference in relapsing patients. Achieved moderate predictive performance (ROC-AUC = 0.72). Highlighted importance of cognitive processes in relapse prediction but limited by small sample and lack of real-time integration.
2021	Kabisa et al.	Statistical analysis (Bivariate and Multivariate Logistic Regression) using retrospective clinical data	Found high relapse prevalence (59.9%), mostly within 2 weeks to 3 months post-discharge. Key predictors included peer influence, family conflict, stress, and polysubstance use. Emphasized environmental and social factors but limited by retrospective design and single-site data.

2.4 Research Gap

Despite the effectiveness of machine learning, psychological, and neurobehavioral approaches in predicting relapse as demonstrated by existing literature, most models lack real-time monitoring capabilities, they are not integrated into deployable clinical systems, limiting their practical application in continuous patient care. Therefore, there is a need for an AI-powered system that supports real-time, continuous, and adaptive relapse prediction.

2.5 Chapter Summary

Although various approaches such as machine learning, deep learning, statistical methods, neuroimaging, and psychosocial analysis have been applied to predict relapse in drugs and substance abuse, significant challenges persist. A major limitation is the reliance on static or pre-treatment data, which fails to capture the dynamic and time-dependent nature of relapse. Additionally, many models focus on single data sources, limiting their ability to represent the complex interaction of behavioral, psychological, and environmental factors. While some studies incorporate longitudinal and real-time data, they are often not integrated into scalable and deployable clinical systems capable of continuous monitoring and timely intervention. Furthermore, traditional statistical approaches struggle to model complex, non-linear relationships compared to modern AI-based techniques. Issues related to generalizability, data availability, and the practicality of certain methods such as neuroimaging also continue to limit real-world application. Addressing these interconnected challenges is essential for developing a comprehensive, integrated, and real-time relapse prediction system that can effectively support clinical decision-making and improve treatment outcomes in substance use disorder management.

Chapter 3: Research Methodology.

3.1 Introduction

This chapter highlights the methodology used in carrying out the study to develop the proposed system. The chapter will include information on the methodology used in the design of the study, methodologies of data collection and analysis carried out during the study, and the methodology used in developing the proposed system. This chapter will also highlight software technology used in the development of the study.

3.2 Research Design

By definition, it can be described as the approach that entails the planning, procedure, and strategy for collecting and analyzing data in the research process. In other words, research design constitutes a set of processes designed to guide a researcher in addressing research objectives through a systematic process in order to get valid, reliable and valuable results from their research (Creswell & Creswell, 2023).

For the purpose of this research, the DSR approach shall be adopted. DSR is a research methodology that has been widely adopted by scholars for research in information systems field for developing and evaluating information system artifacts. The purpose of DSR is to create innovative artifacts such as model, system or framework to solve a given problem in the real world environment while maintaining rigor and relevance (Gregor & Hevner, 2021). In general, there are different phases involved in DSR methodology and include; problem identification and motivation, definition of solution objectives, design and development, demonstration, evaluation, and communication (Peffers et al., 2020).

In relation to this specific research problem, the problem to be solved in the study is that there is no relapse prediction tool in FASAA's patient population. The goal of this solution is to develop an artificial intelligence-based relapse prediction system for use among FASAA's patients. The design and development of such system will be done through utilizing various technologies like Django, React, Logistic Regression and Random Forest. The step of demonstration entails applying the system to patients' data while evaluation will focus on checking the accuracy of the model developed. Lastly, documentation of the whole process will entail the step of communication.

The adoption of the above research methodology for designing the relapse prediction tool in FASAA's is due to the fact that there exists a need to have a robust system that will predict the relapse problem at hand.

3.2.1 Data Collection Approaches

This study employed the mixed method approach for the purpose of collecting qualitative as well as quantitative data more easily. The primary data collection method used included the use of structured questionnaires and interviews, which provided useful information on the existing methods for predicting relapse. Secondary data was gathered using the review of previous literature, records of patients, and other past studies, which enabled the researcher to identify both past and present relapse prediction methods used in the case of drug and substance abuse. Thus, through the use of these approaches in gathering data, the objectives of the research were successfully achieved.

3.3 Population and Sampling

3.3.1 Population

As noted by Willie (2023),a population is defined as all people or entities who possess certain attributes of interest to the researcher as subjects of investigation. In this case, the population of interest in this research comprises drug and substance abuse patients and their caregivers at FASAA. Nevertheless, due to the limitations of time and other constraints, sample selection will be necessary.

3.3.2 Sampling

This refers to the procedure where a smaller number of people called a sample is selected from a larger group called a population to participate in the research.

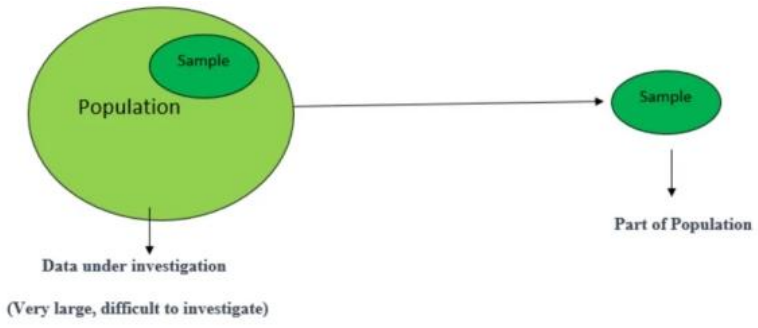


Figure 1: Sampling Example

3.3.3 Justification of Sampling Technique

This study used purposive sampling to select participants from FASAA Rehabilitation Center. This technique involves deliberately choosing individuals with relevant knowledge and experience related to the research objectives. It was appropriate because the study focuses on developing a relapse prediction system, which requires input from individuals directly involved in rehabilitation processes and patient management.

The participants included social workers, counselors, psychologists, nurses, and administrators who are actively engaged in patient monitoring, treatment planning, and relapse prevention. These participants were selected because they can provide accurate and relevant information about current practices and system challenges.

Purposive sampling ensured the collection of high-quality, meaningful data by focusing on knowledgeable respondents, thereby supporting the effective design of the proposed system and improving relapse prediction outcomes.

Sample	**Actual Size**
Patients	20
Health Care Professionals	10
Total	**30**

Table 3.1 Target Population

3.4 Research Instruments

These refer to the tools that help researchers obtain useful information from their subjects in order to accomplish the purposes of their studies. In this study, the tools used in collecting data included semi-structured interviews and questionnaires. This helped the researcher obtain both quantitative and qualitative data concerning the management of drug and substance abuse cases and the need for an AI-driven relapse prediction tool.

3.4.1 Questionnaires

Questionnaire refers to a survey research instrument which comprises of pre-designed questions aimed at collecting data to gain insight regarding respondent's knowledge and opinions (Caduff, 2023). The questions may either be open-ended or closed-ended. In this instance, the questionnaire was adopted to obtain information about the prevailing relapse prediction methods, the problems encountered in the process, and their effectiveness. It was also used to gather opinions regarding the suggested system of using artificial intelligence to predict relapse and the advantages associated with it.

3.4.2 Interviews

Interview is a method of collecting data through direct communication between the researcher and the subjects under study. This method was used in this study where semi-structured interviews were conducted involving various healthcare professionals like psychologists, nurses and other patients.

3.5 Requirements Analysis

In this chapter of the paper, it will be necessary to analyze the requirements for the system under consideration on the basis of functions and features. Specifically, the section will discuss the identification of the major users, their interactions, and challenges connected with patient behavior monitoring and relapse prediction.

3.5.1 Functional Requirements

As we have said before, functional requirements relate to what should be done by the system.

i. The system must allow the Administrator to register patients, staff members and assign patients to specific social workers.
ii. The system must allow patients to log into their accounts securely, complete daily check-in forms, interact with an AI chatbot for emotional support, write and submit daily journal entries.
iii. The system must store all patient data (check-ins, chatbot conversations, and journals) in a database.
iv. The system must analyze patient data using an AI model to predict relapse risk, generate a relapse risk score based on patient input and automatically send alerts to the assigned social worker when a high-risk condition is detected.

v. The system must allow social workers to view patient data and risk assessments, provide intervention and update patient records.

3.5.2 Non-Functional Requirements

It refers to all those properties and attributes associated with the entire system.The system must be reliable and consistently available for users.

i. The system must provide accurate AI predictions based on patient data.
ii. The system must ensure data security and protect sensitive patient information.
iii. The system must have a user-friendly and responsive interface.
iv. The system must be scalable to support multiple users simultaneously and perform efficiently with minimal delays during data processing and AI analysis.

3.5.3 Software Requirements

i. Operating System: Windows 10 or Windows 11
ii. Backend Framework: Django (Python)
iii. Frontend Framework: React.js
iv. Database: PostgreSQL
v. Programming Languages: Python, JavaScript
vi. Development Tools: Visual Studio Code, Jupyter Notebook

3.5.4 Hardware Requirements

i. Computer or laptop capable of running development tools
ii. Smartphones or computers for system users (patients and staff)
iii. Stable internet connection for accessing the web-based system

3.6 System Development Process

This section discusses the tools and technologies utilized in the development of the prototype of the system. The selection of these tools was based on their capability to enable the design and creation of an efficient and scalable system with the use of artificial intelligence.

3.6.1 System Development Methodology

System development methodologies are the approaches that can be applied when designing, developing and testing the systems in question. The process is broken down into several steps to ensure efficient execution.

In this research work, the evolutionary prototyping methodology will be applied. According to this method, the system is being developed step-by-step and gradually improved through testing and refinement.

The methodology is best suited for this particular research project as it helps integrate complex components like the AI model.

3.6.2 Evolutionary Prototyping

Evolutionary prototyping involves developing a working version of the system with limited functionality, which is gradually improved over time. The prototype serves as an early model of the system, allowing users and stakeholders to interact with it and provide feedback.

This approach enables the researcher to better understand user requirements and identify any missing features or improvements needed before final implementation. Even though the first prototype may not capture all aspects of the system, it still holds significant importance in the designing process of the system.

Through continuous testing, evaluation, and enhancement, the prototype evolves into a fully functional system. This method is particularly effective for this project as it supports iterative development of both the system features and the AI-based relapse prediction component.

3.7 Summary of how the system works

The proposed system is a web-based application that uses random forest and logistic regression ML algorithms to predict the likelihood of relapse in DSA patients. The system operates by collecting patient data through three primary sources: daily check-in forms, AI chatbot interactions, and journal entries. These inputs provide both structured and unstructured data reflecting the patient's emotional, psychological, and behavioral state which is essential to predict relapse risk in DSA patients.

3.7.1 Patient Module

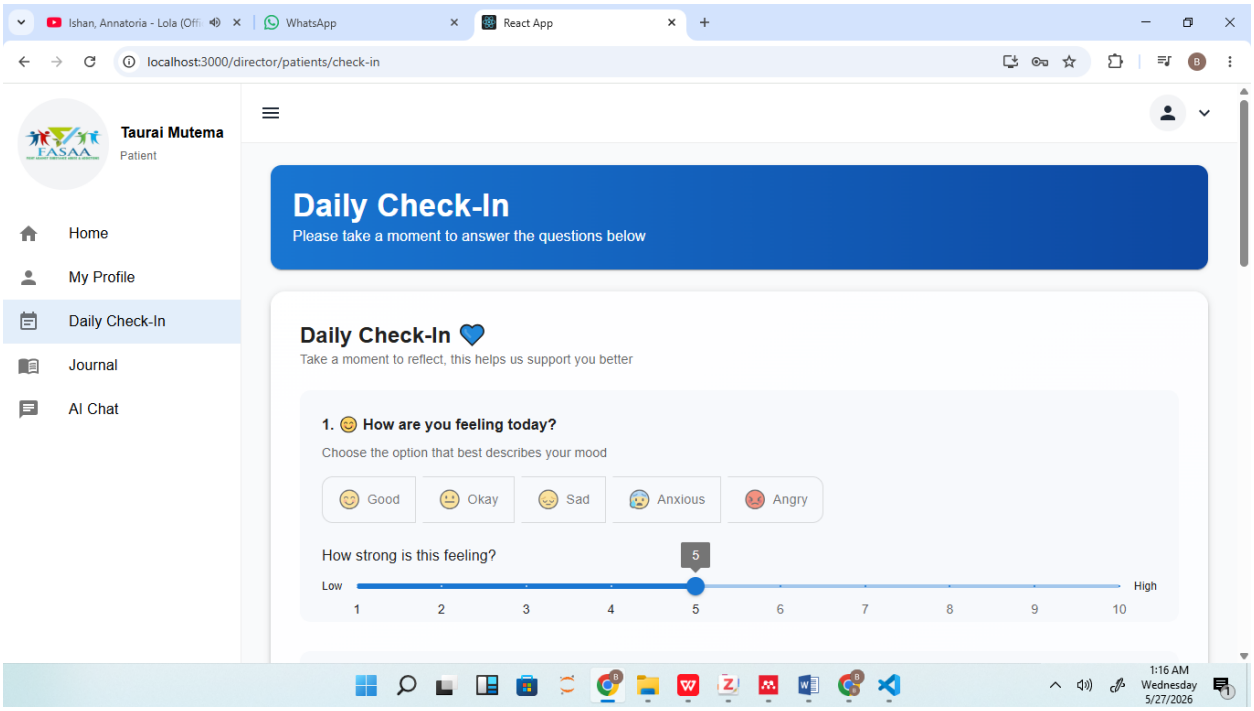


Figure 2: Patient Daily Check-in form

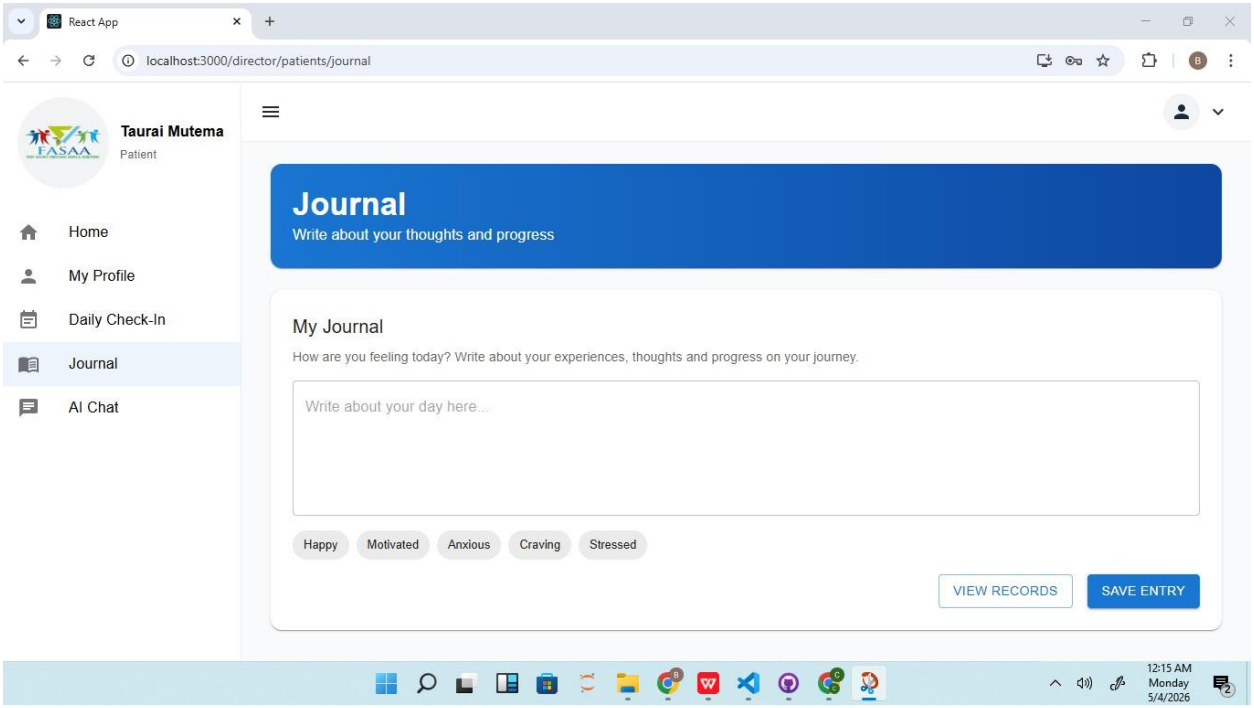


Figure 3: Patients journal for writing down their experiences, thoughts and progress

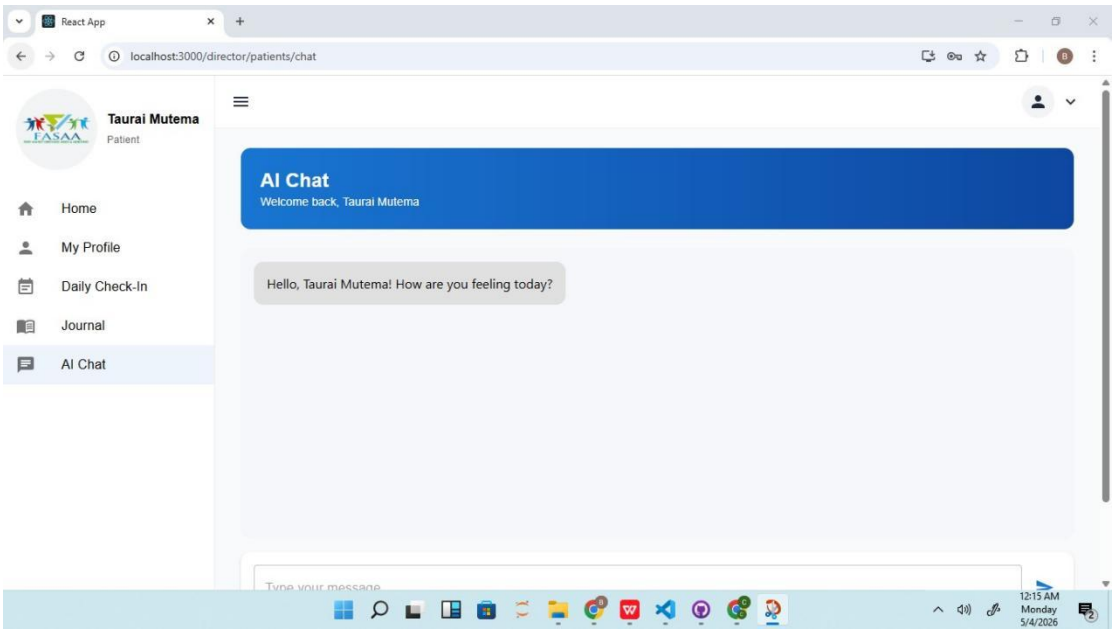


Figure 4: Patients AI chatbot

All collected data through patient check in form, journal and chatbot is processed by an AI-based relapse prediction model. The model analyzes patterns in the patient's responses and generates a risk score, which is used to classify the patient into categories such as low, moderate, or high risk of relapse. The system then automatically generates an alert showing relapse risk and reasons and sends a notification to the assigned Social Worker. The Social Worker can then review the patient's data and take appropriate action, such as counselling or further monitoring.

3.7.2 Social Worker Module

A particular social worker is allocated to each patient in the system. Every time there is an alert, it automatically appears on the dashboard of the allocated social worker, ensuring timely attention and follow-up.

Two machine learning algorithms have been used in predicting relapses. The first algorithm uses the Random Forest method of training and is linked to the patient Check-In Form. Alerts from this algorithm will appear on the Alerts & Notifications page.

The other algorithm utilizes the Logistic Regression technique of training and is integrated with the Chatbot and Journal modules. Alerts from this algorithm will appear on the Relapse Risk Monitor page, allowing the monitoring and follow-up of any patient with a likelihood of developing relapse.

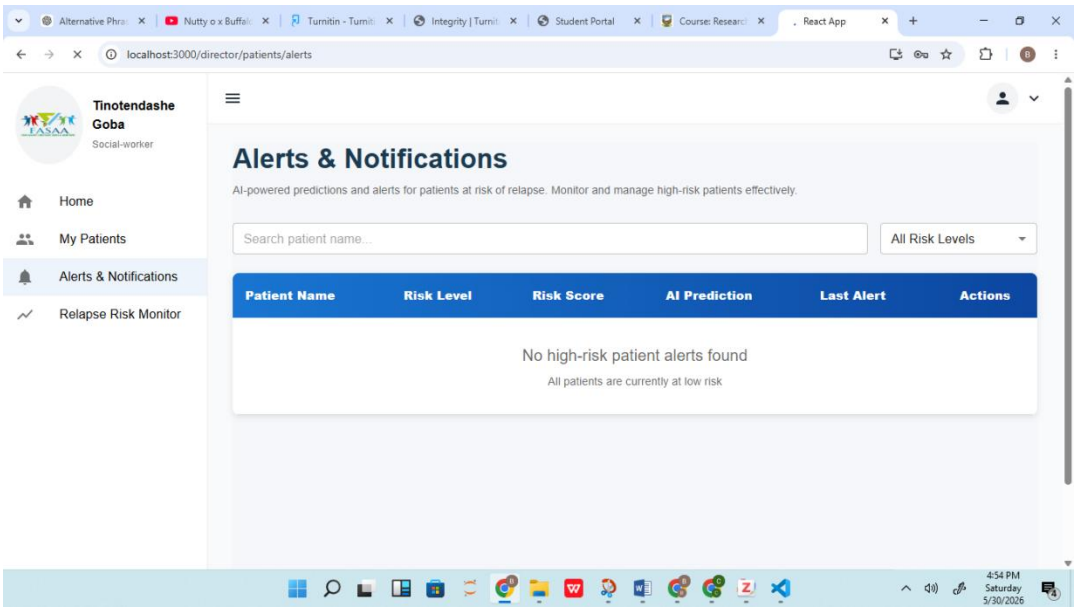


Fig 5: Screen showing the relapse and alerts notifications generated from the patient check-in form

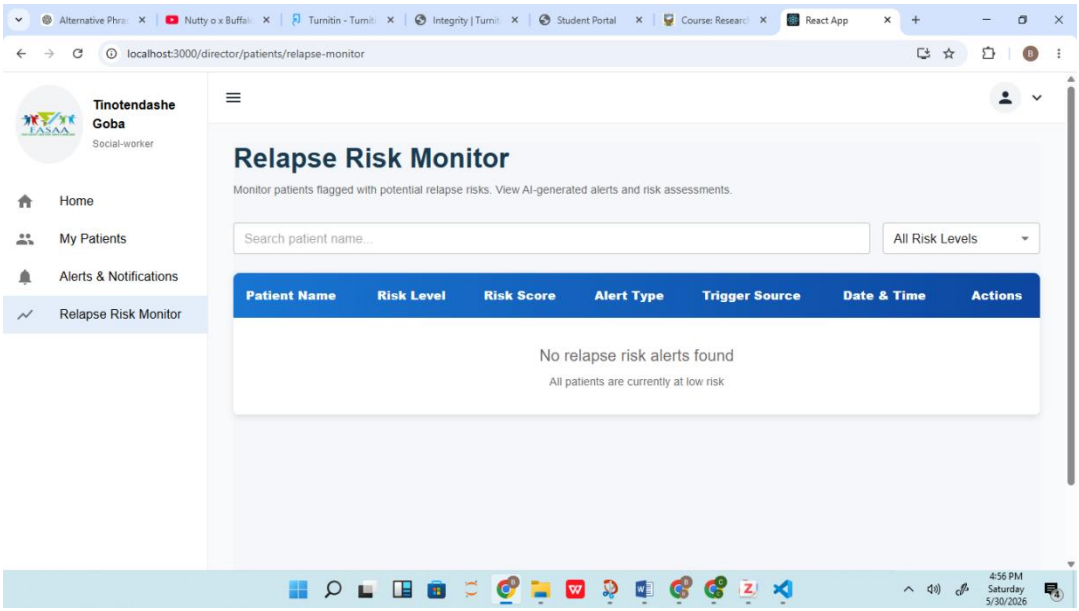


Figure 6: Screen showing relapse risk monitor where notifications from chatbot and journal are shown

3.8 AI Model Training (Jupyter Notebook)

The AI model used in this system was developed and trained using a Jupyter Notebook environment. The training process began with the creation of a simulated dataset that reflects real-world patient check-in data. The dataset included key behavioral and psychological features such

as mood, mood intensity, stress level, sleep hours, energy level, craving strength, trigger count, isolation status, number of positive actions, and confidence level.

3.8.1 Training the Random Forest Relapse Prediction Model used in the Daily Check-in Form

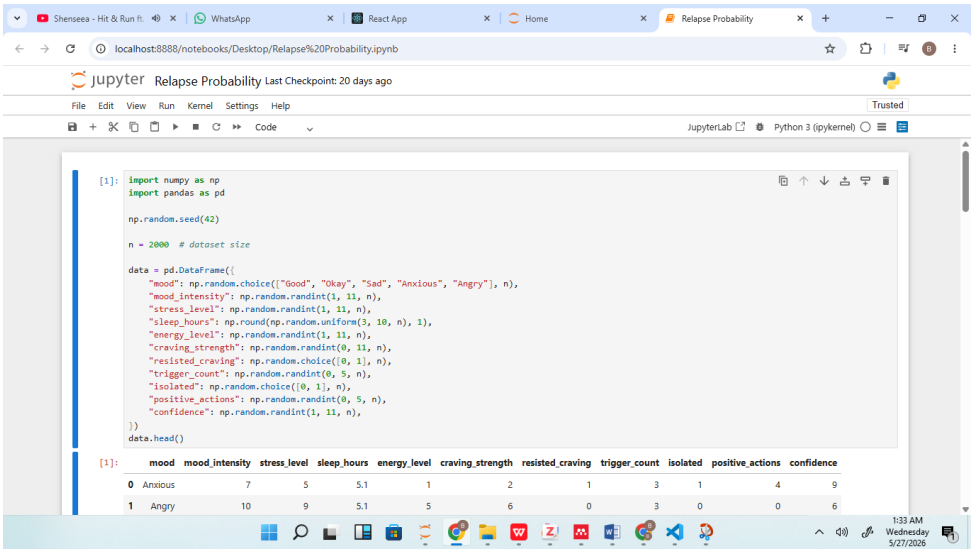


Fig 7: Sample from Simulated Dataset for Model Training

In order to make the dataset realistic and diverse, randomness was added while generating the data set. This approach allowed the model to learn general patterns rather than memorizing fixed values, thereby reducing overfitting and improving prediction reliability.

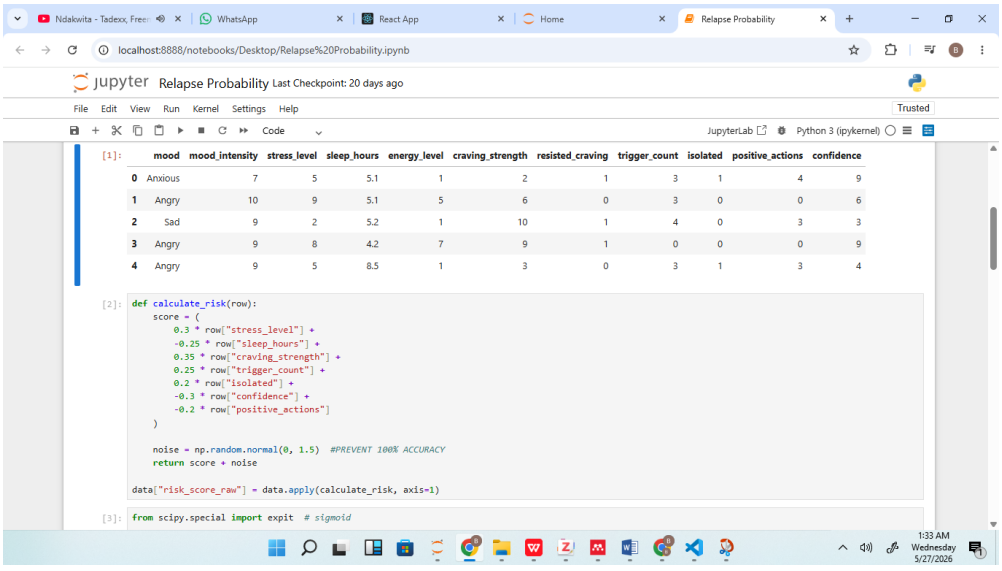


Fig 8: Code Snippet Showing Dataset Simulation with Randomization

After generating the dataset, preprocessing was performed. Categorical variables such as mood were converted into numerical representations using one-hot encoding, while multi-select inputs such as triggers and positive actions were transformed into numerical counts. After that, the dataset was split into train and test parts to test the accuracy of the model. Random Forest Classifier was chosen as the main algorithm since it is a robust method and can capture the relationships among behaviors.

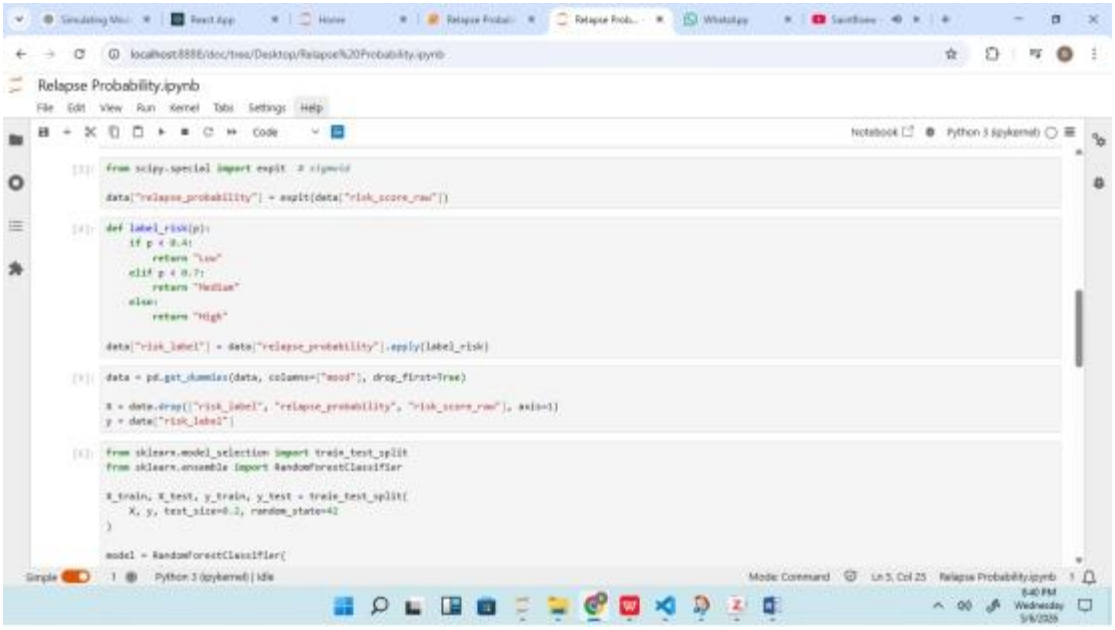


Fig 9: Data preprocessing and encoding features, Model training using Random Forest Classifier

To generate meaningful predictions, a probability-based approach was implemented. Instead of producing binary outputs, the model calculates the likelihood of relapse using probability scores, which are then converted into percentage values and categorized into Low, Medium, and High risk levels. Model evaluation was conducted using accuracy as a performance metric. The model gave reasonable accuracy results, meaning it generalizes reasonably well without overfitting. Lastly, the model was exported into a pickle file named relapse_model.pkl together with its respective feature column in a separate file known as model_columns.pkl, and both files imported into the django backend for real-time predictions.

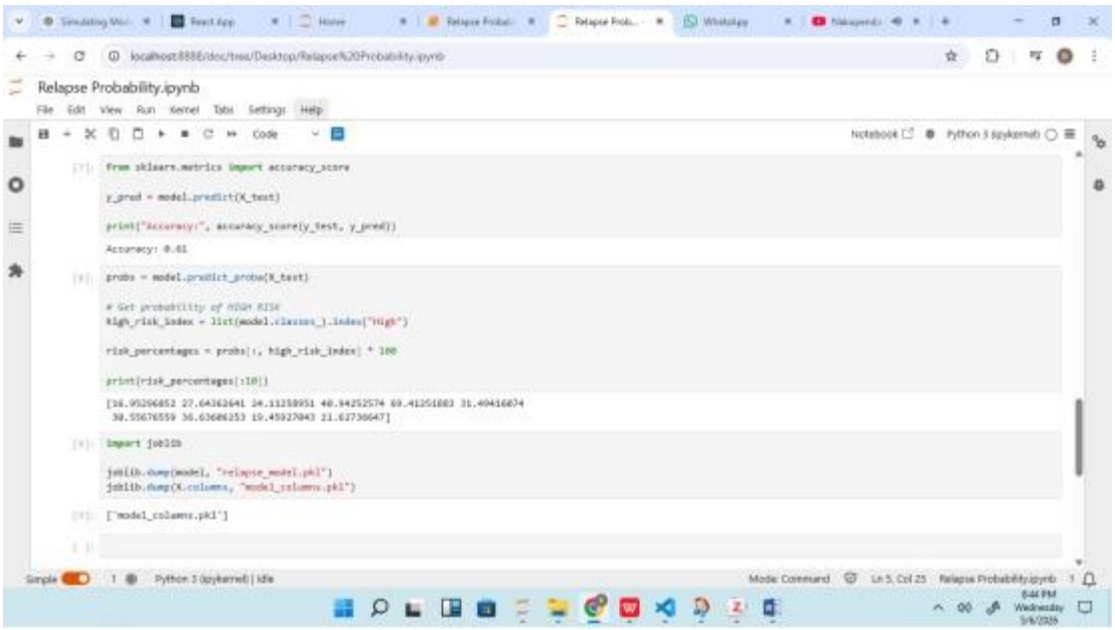


Fig 10: Probability-Based Prediction Output from the Model, Model Accuracy Evaluation Result and Saving the Trained Model and Feature Columns

3.8.2 Training the Machine Learning model used in Chatbot and Journal

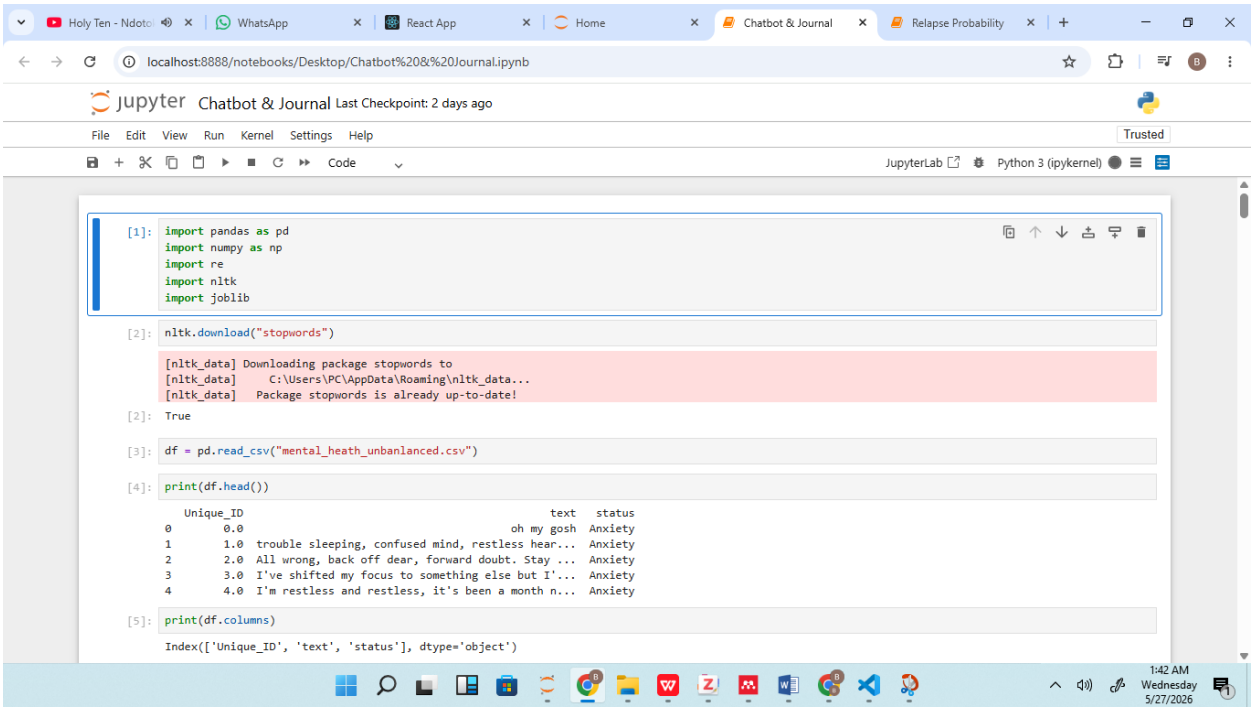


Fig 11: Model Training

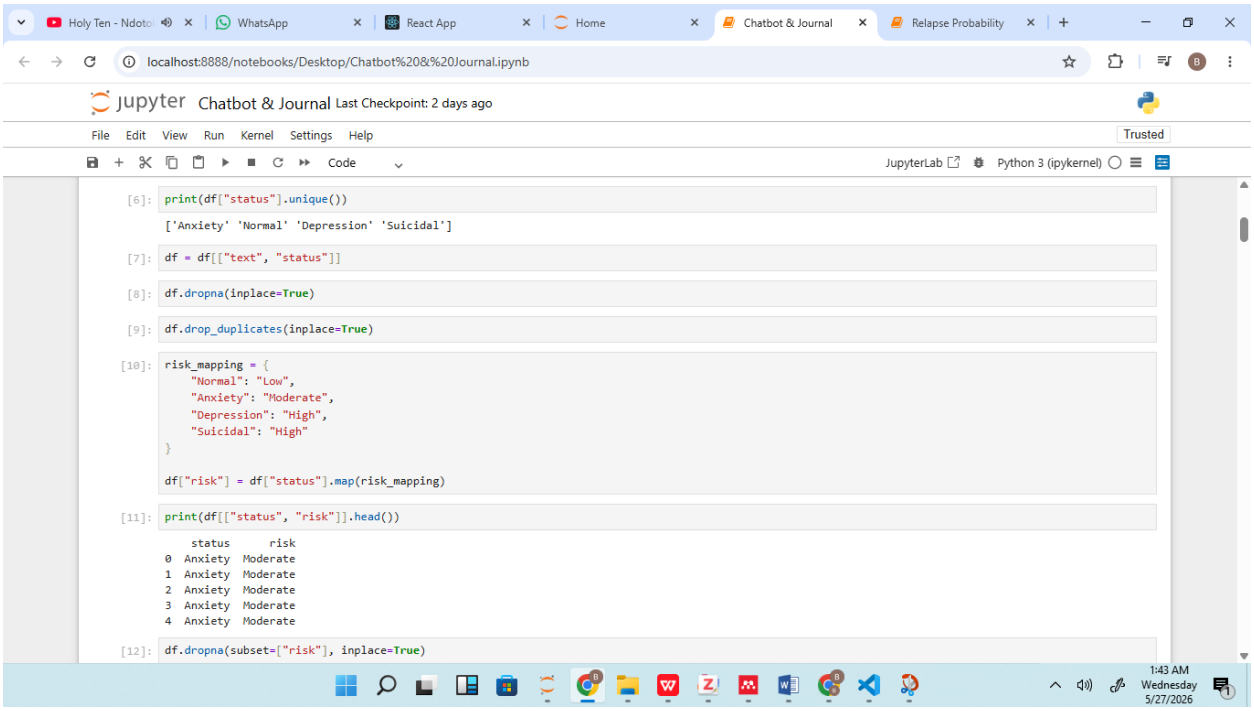


Fig 12: Model Training

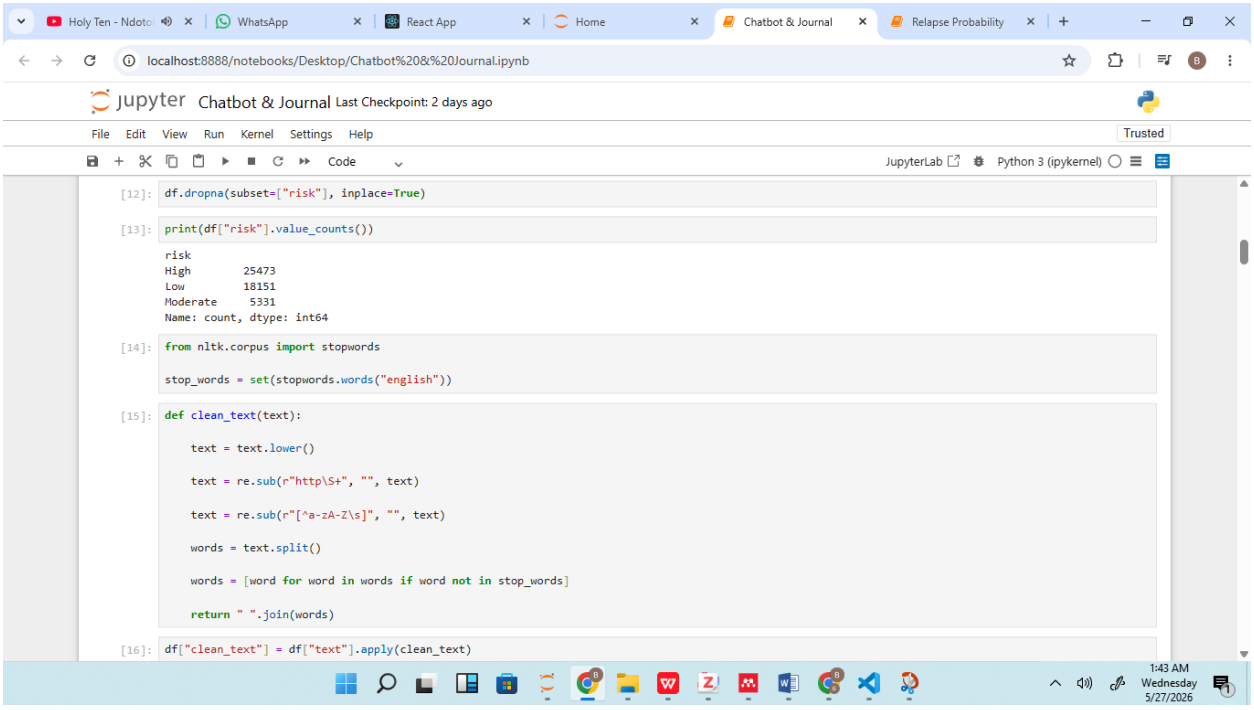


Fig 13: Model Training

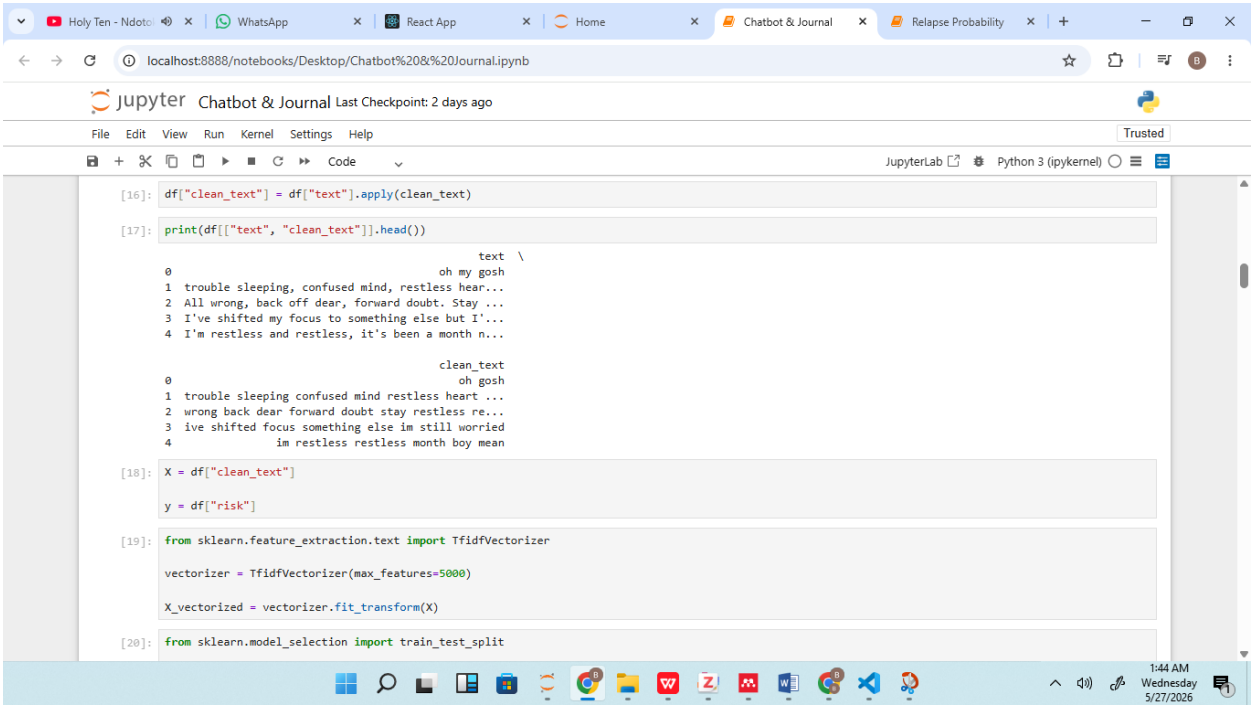


Fig 14: Model Training

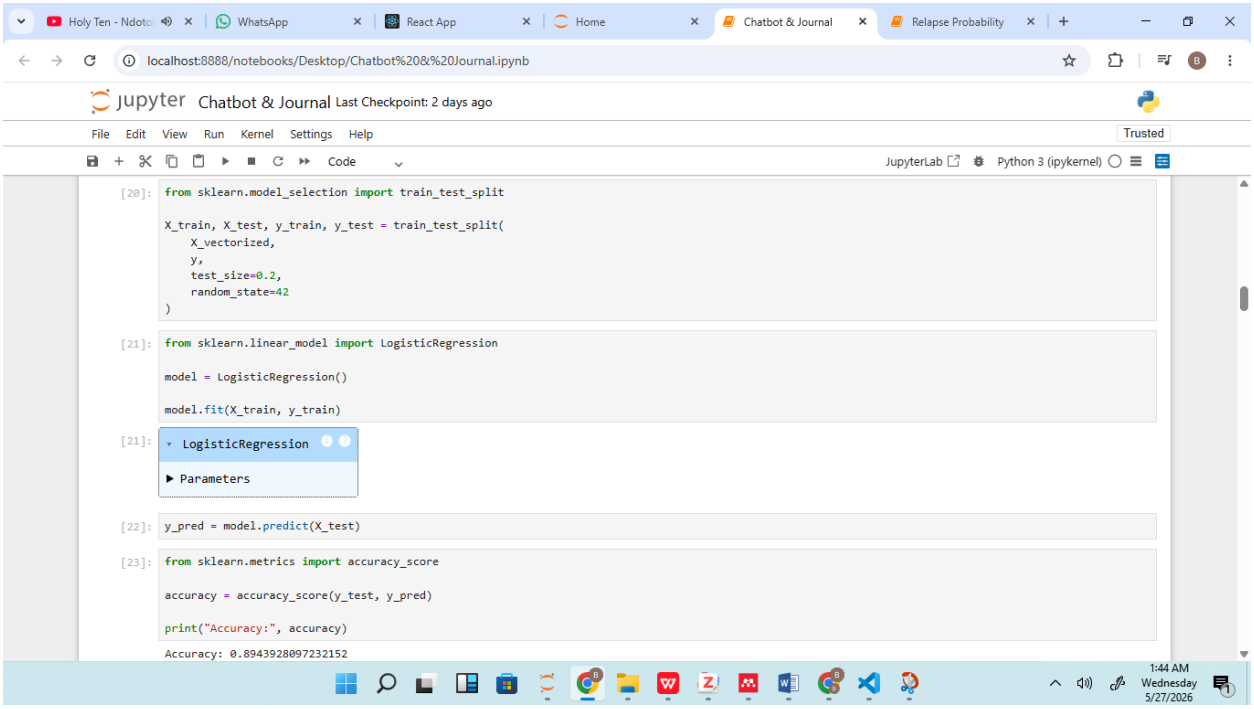


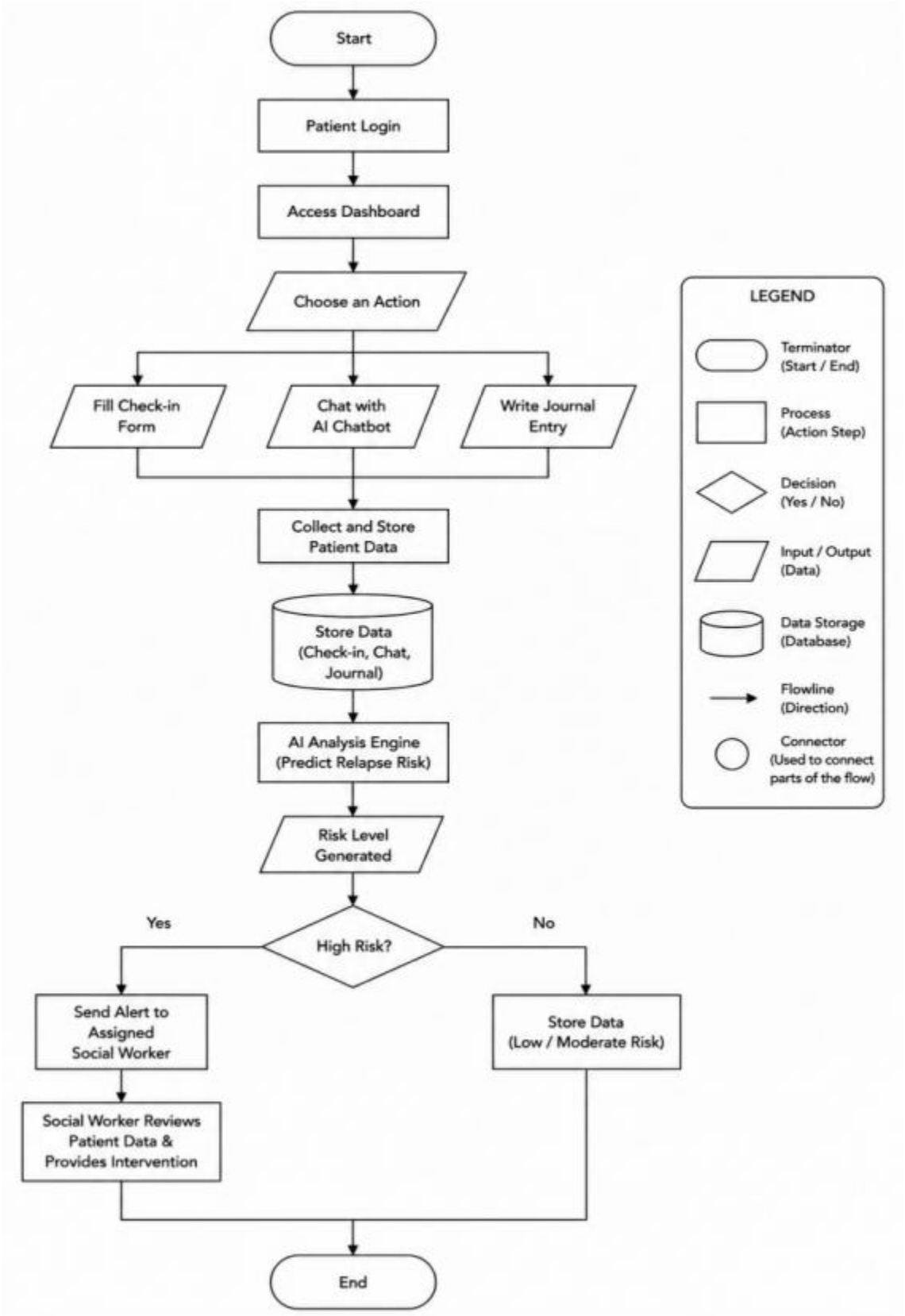
Fig 15: Model Training

3.9 System design

This stage outlines how the system's components and data structures are designed to meet and fulfill the specified requirements.

3.9.1 Proposed System Flowchart

In essence, the flowchart is a visual way of showing the steps that the system follows to execute its processes based on the view of the patient. In the system, first, the patient logs into the system and goes to his or her dashboard. Then, the patient may do one of the following: fill out a check-in form, engage in dialogue with the AI chatbot, or create a journal entry. All collected information is then analyzed by the AI model. After evaluating the data, the AI model assesses the risk of the patient having a relapse. If the risk is high, an alert is sent to the assigned Social Worker. If the risk is low or moderate, the data is simply stored for monitoring purposes. This flowchart is useful for understanding the logical sequence of operations and how decisions are made within the system.


Start
Patient Login
Access Dashboard
Choose an Action
Fill Check-in Form
Chat with AI Chatbot
Write Journal Entry
Collect and Store Patient Data
Store Data (Check-in, Chat, Journal)
AI Analysis Engine (Predict Relapse Risk)
Risk Level Generated
High Risk?
Yes
No
Send Alert to Assigned Social Worker
Social Worker Reviews Patient Data & Provides Intervention
Store Data (Low / Moderate Risk)
End
LEGEND
Terminator (Start / End)
Process (Action Step)
Decision (Yes / No)
Input / Output (Data)
Data Storage (Database)
Flowline (Direction)
Connector (Used to connect parts of the flow)

3.10 Testing

Testing plays an important role in the system development process, which consists of evaluating the quality, functionality, reliability, and performance of a software application to confirm that it meets the required standards and users' needs. The evaluation process aims to identify any errors or deficiencies within the system before releasing it to guarantee its optimal operation.

3.10.1 Black Box Testing

It is testing technique that examines functionality, usability, performance, and reliability of a system in accordance with the requirements and demands of its end user without analyzing its internal processes. Black-box testing takes its name from the fact that just like a black box, the internal operations of a computer system cannot be observed by the tester who tests its functioning. The technique is commonly applied when developing software applications to ensure that they meet users' needs (Pressman & Maxim, 2020).

3.10.1.1 Testing the Patient Module

The Patient Module is designed for patients recovering from drug and substance abuse. Through this module, patients can securely log into the system and access various support features. Patients are able to complete daily or periodic check-in forms that capture information related to their emotional state, behaviour, cravings, stress levels, and recovery progress. The module also includes a journaling feature that allows patients to record their thoughts, feelings, and daily experiences during recovery. Furthermore, patients can interact with an AI-powered chatbot that provides emotional support, guidance, and basic recovery-related assistance. Information collected through the check-in forms, journals, and chatbot interactions is analysed by the AI relapse prediction model to determine the patient's relapse risk level.

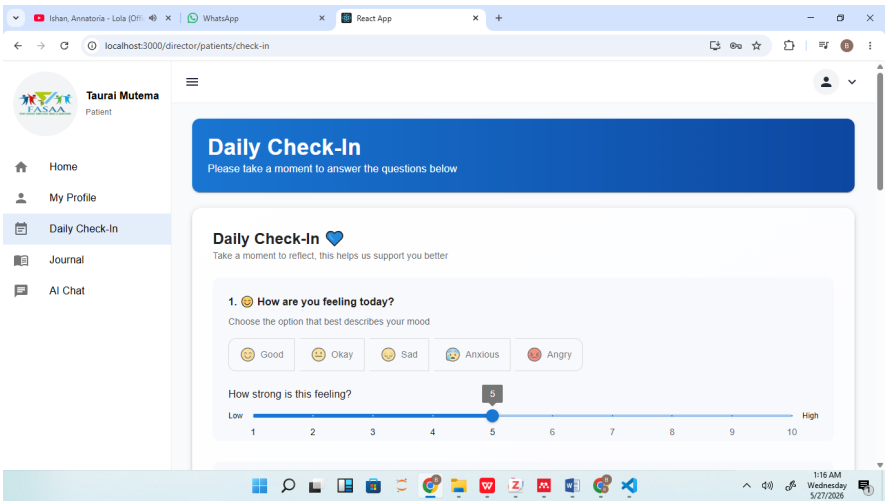


Fig 15: Patient Checkin

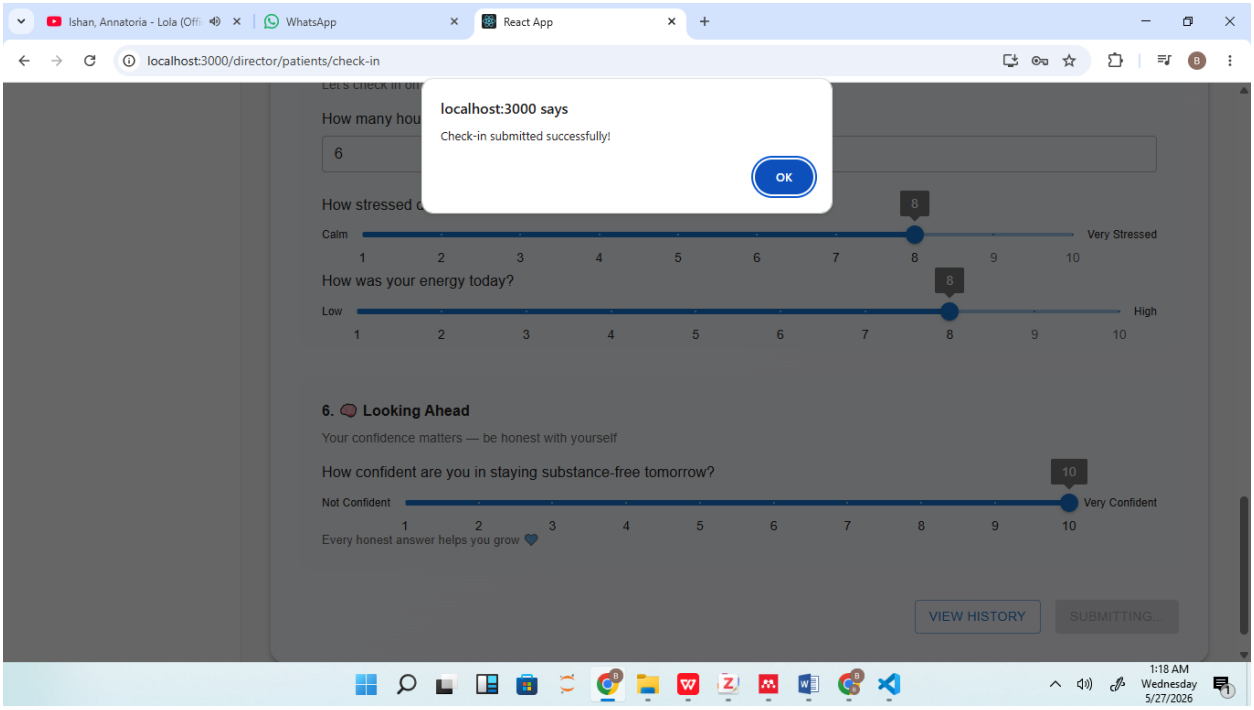


Fig 16: Checkin submitted

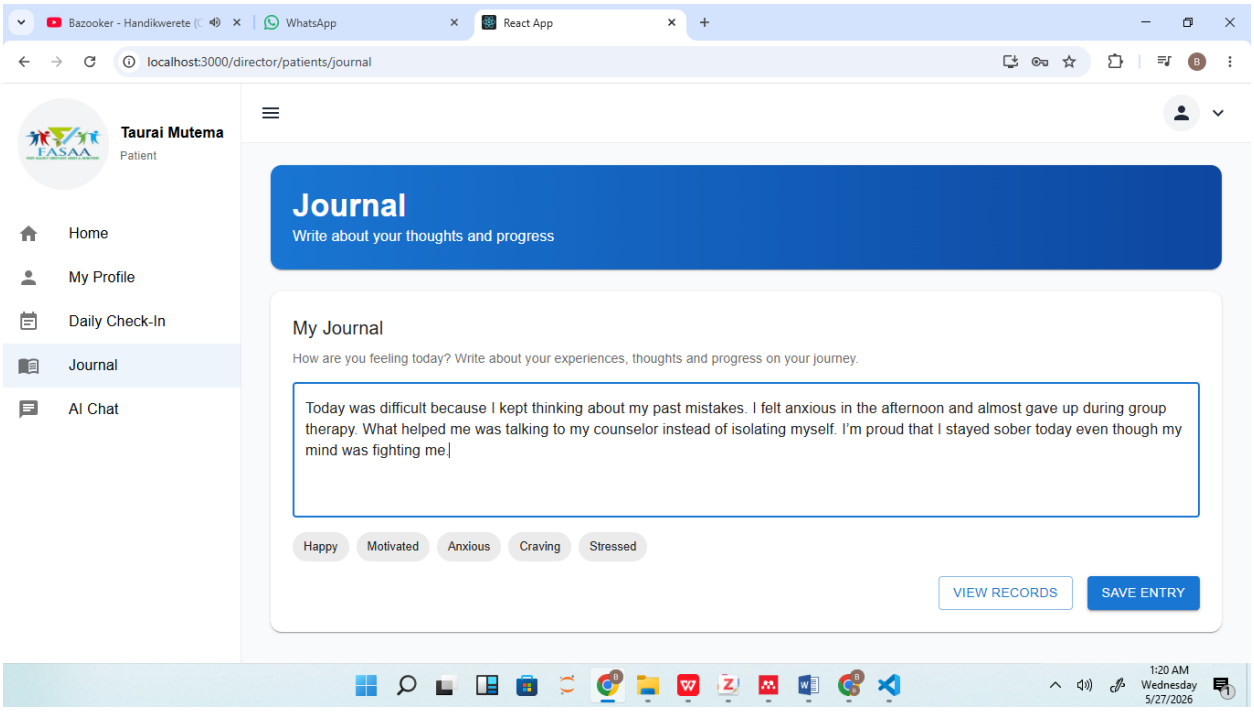


Fig 17: Writing the journal

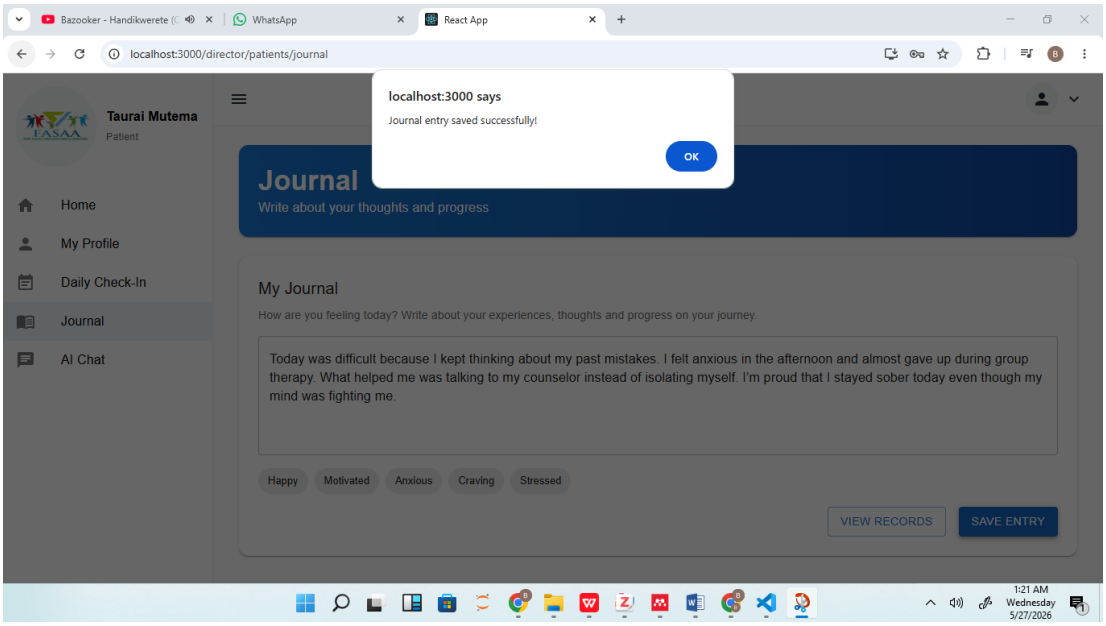


Fig 18: Journal successfully saved

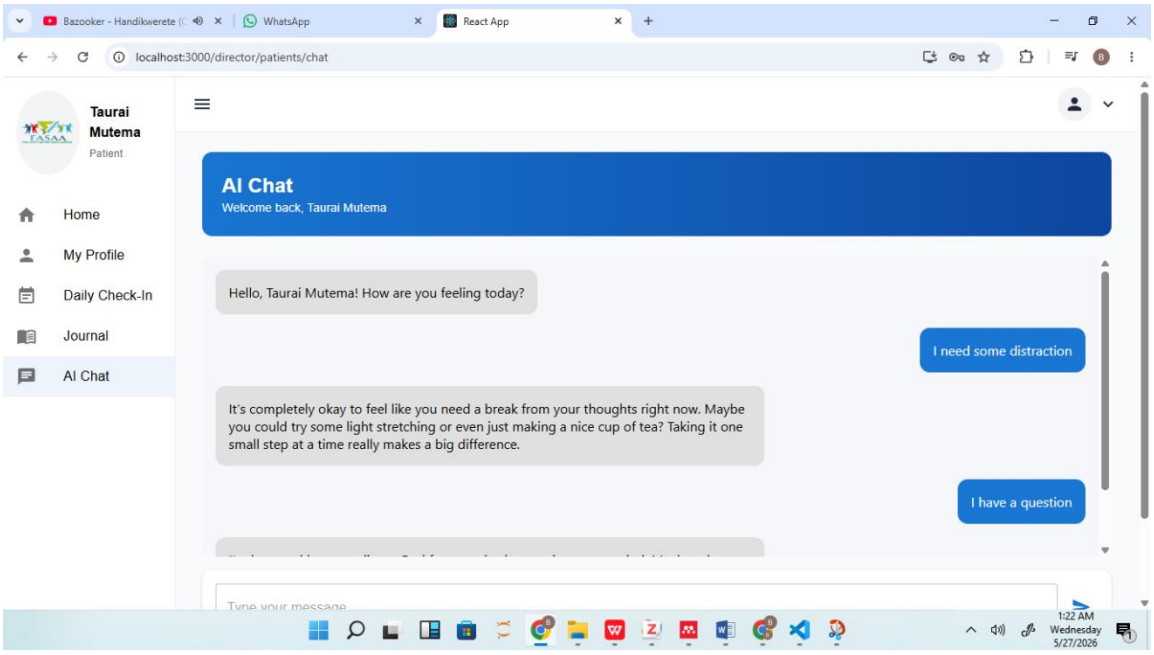


Fig 19: Chat with chatbot

3.10.1.2 Testing the Social Worker Module

The Social Worker Module is intended for social workers responsible for monitoring and supporting assigned patients. Through this module, social workers can access information about their assigned patients and receive alerts generated by the AI relapse prediction system whenever a patient is identified as being at risk of relapse. The alerts help social workers to quickly identify patients who may require immediate intervention or additional support. Social workers can review patient check-in reports, journal entries, and risk assessments in order to better understand the patient's condition and determine suitable recovery strategies or counselling actions. This module improves communication, monitoring, and timely intervention in the patient recovery process.

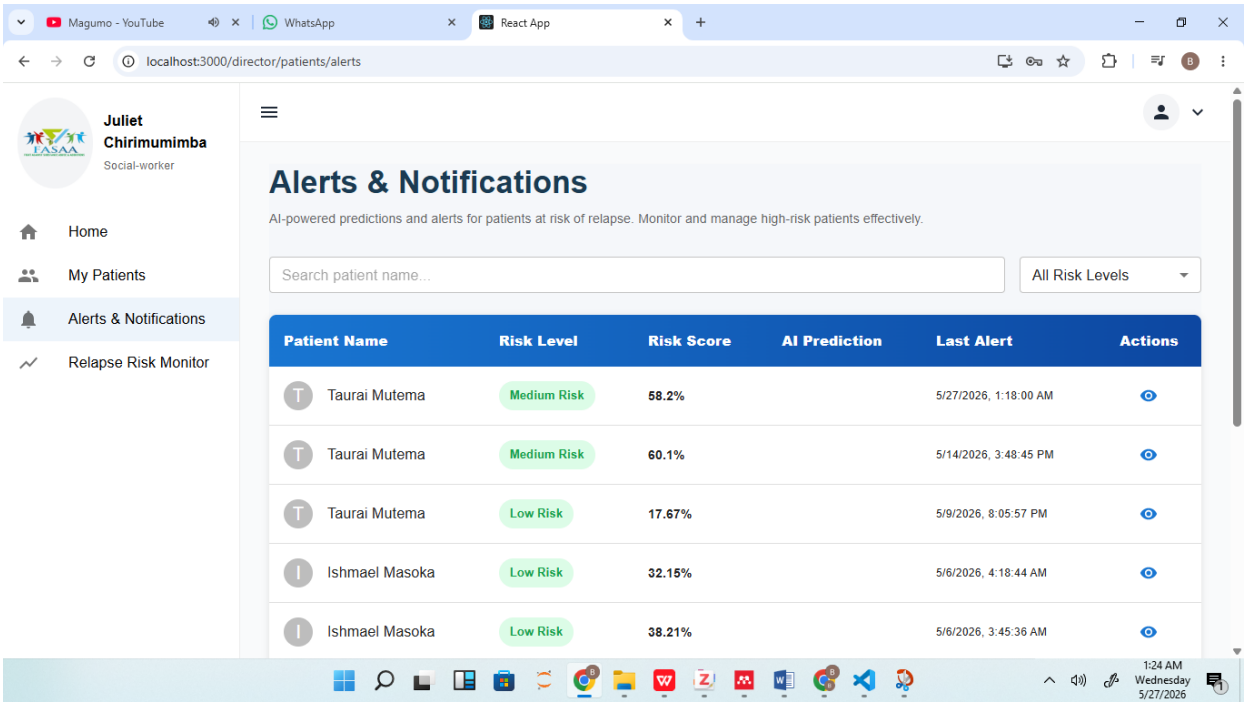


Fig 20: Alert and Notifications from the Check In form

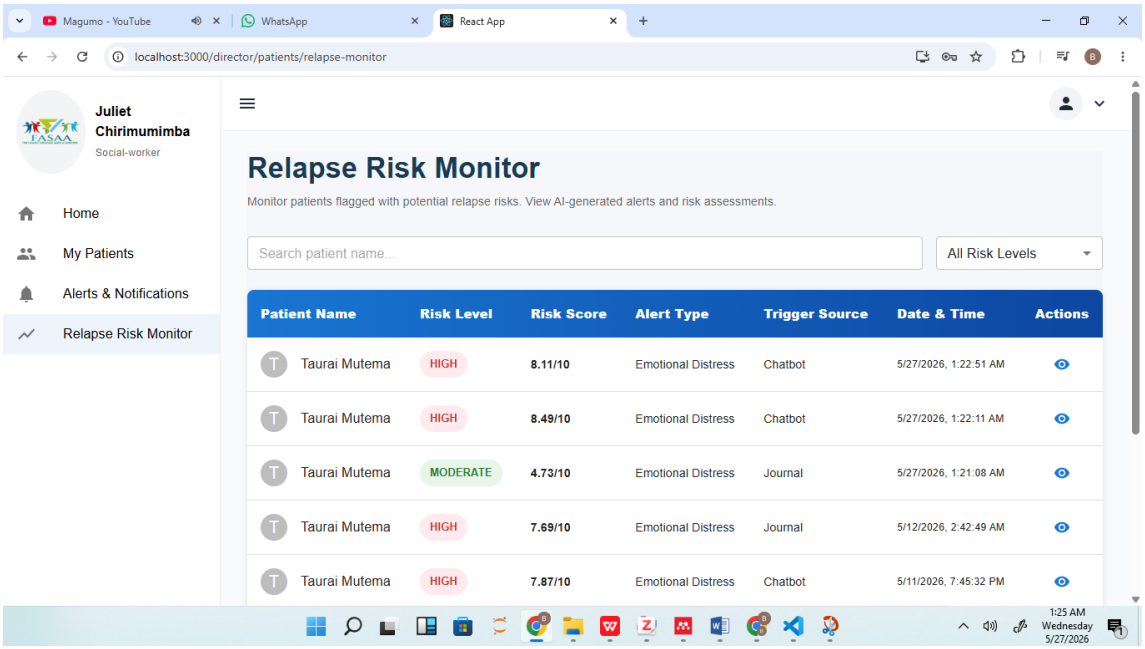


Fig 21: Relapse Risk Monitor From Jornals and Chatbots

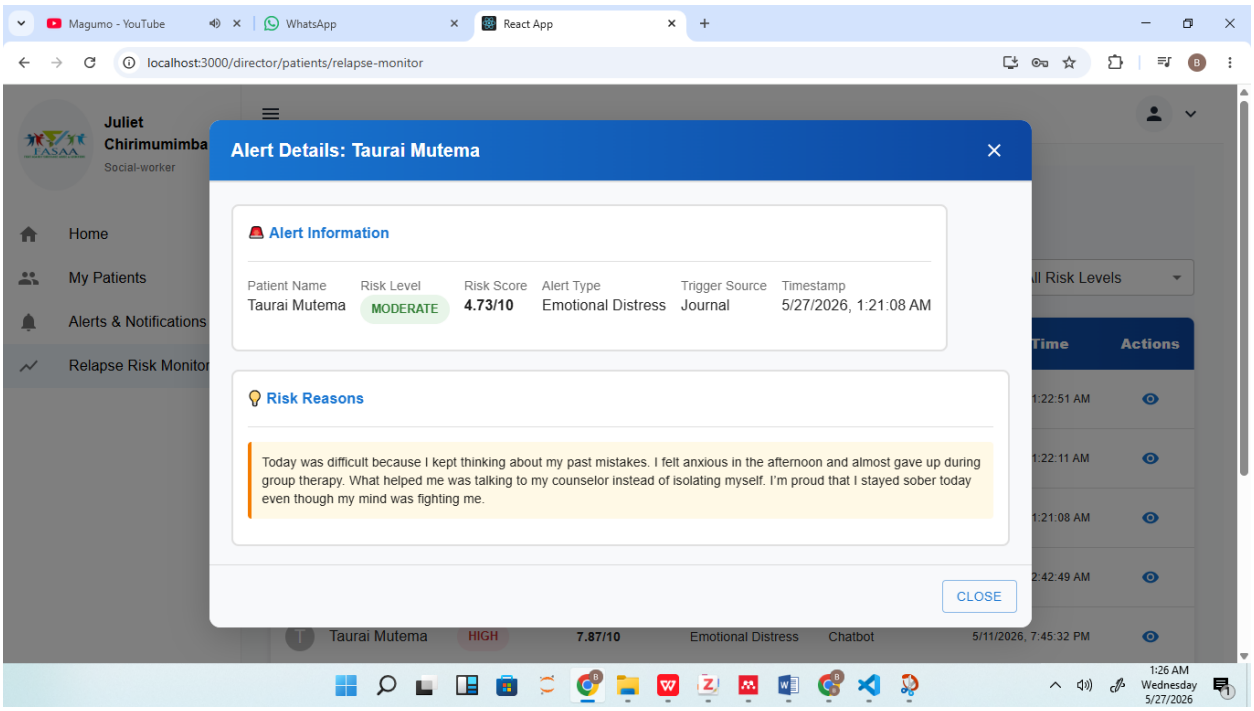


Fig 22: Journal Alerts Details

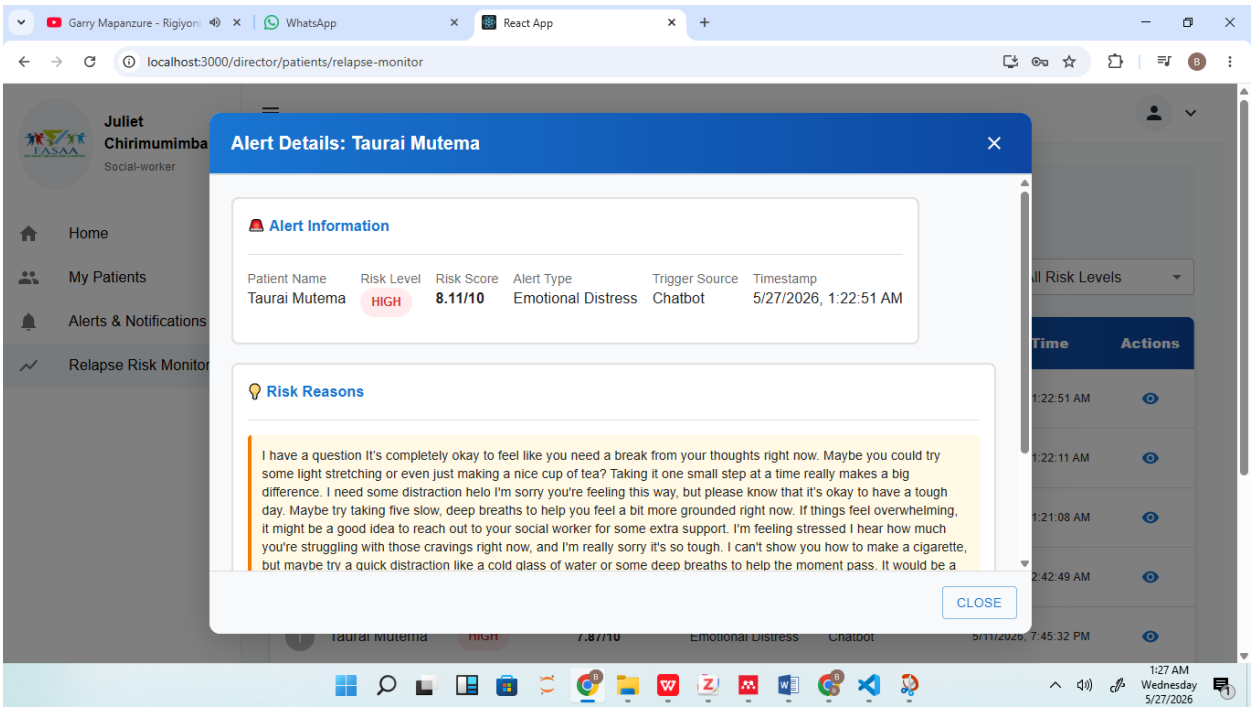


Fig 23: Chatbot Alerts Details

3.11 Chapter Summary

In this chapter, discussion of the research methodology adopted in the research has been carried out. Specifically, in this chapter, discussion about the research methodology employed to develop the AI-based system for predicting relapse among patients under substance abuse and drug abuse was carried out. The research methodology involved both mixed method where data was gathered qualitatively and quantitatively. Primary data was collected using questionnaires and interviews, whereas secondary data was collected using literature and documented sources. Purposeful sampling method was applied to select qualified participants in the study from FASAA Rehabilitation Center. The process of analyzing the collected data and its implication to the development of the AI-based prediction system was also discussed in this chapter.

Chapter 4 - Data Presentation, Analysis and Interpretation

4.1 Introduction

In this chapter, analysis and interpretation of data for the research on the design of a relapse prediction model at the FASAA rehabilitation center will be carried out. The purpose of this research project is to examine the effectiveness of the system in improving patient care through prediction of relapse cases. From this chapter, the effectiveness of the system in aiding relapse prediction using Artificial Intelligence, and assisting through its various components such as the chatbot and journaling application will be evaluated. This chapter forms the basis for the conclusion and recommendation that will be made in the next chapter.

4.2 Evaluation Measures and Metrix

Accuracy, Precision, Recall, Confusion Matrix and F1 Score are some of the metrics to be used to evaluate the model's performance in relapse prediction.

4.2.1 Evaluating the Random Forest Relapse Prediction Model used in the Daily Check-in Form

The relapse prediction model was developed using Machine Learning Classification algorithms on the Jupyter Notebook. Below are the results for model performance in terms of the confusion matrix and classification report.

```
Confusion Matrix:
[[127  29   0]
 [ 27 117   0]
 [ 54  46   0]]

Accuracy: 0.61
Precision: 0.4575
Recall: 0.61
F1 Score: 0.5228571428571428

Classification Report:
              precision    recall  f1-score   support

        High       0.61      0.81      0.70       156
         Low       0.61      0.81      0.70       144
      Medium       0.00      0.00      0.00       100

    accuracy                           0.61       400
   macro avg       0.41      0.54      0.46       400
weighted avg       0.46      0.61      0.52       400
```

Figure 24: Confusion Matrix and Classification Report Results of the AI-Powered Relapse Prediction Model Generated in Jupyter Notebook.

Accuracy Calculation

The accuracy of the model was calculated to determine the overall percentage of correctly classified predictions.

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}}$$

From the confusion matrix:

Correct Predictions = 127+117+0=244

Total Predictions=400

Therefore:

$$\text{Accuracy} = \frac{244}{400}$$

=0.61

=61%

The model achieved an accuracy score of 61%, meaning that 244 out of 400 patient cases were correctly classified.

Precision Calculation

Precision was used to measure how many predicted positive cases were actually correct.

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

For the High-risk class:

TP=127

FP=27+54=81

Therefore:

$$\text{Precision} = \frac{127+81}{127}$$

$$= \frac{208}{127}$$

=0.61

The precision score of 61% indicates that the model correctly identified 61% of the patients predicted as high-risk.

Recall Calculation

Recall was calculated to determine how well the model identified actual positive cases.

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

For the High-risk class:

TP=127

FN=29

Therefore:

$$\text{Recall} = \frac{127+29}{127}$$

$$= \frac{156}{127}$$

$$= 0.81$$

The recall score of 81% shows that the model successfully identified most of the actual high-risk patient cases.

F1-Score Calculation

The F1-score was used to evaluate the balance between precision and recall.

$$F1 = 2 \times \frac{\text{Precision} + \text{Recall}}{\text{Precision} \times \text{Recall}}$$

Using the obtained precision and recall values:

Precision=0.61

Recall=0.81

Therefore: $F1 = 2 \times \frac{0.61+0.81}{0.61 \times 0.81}$

$= 2 \times \frac{1.42}{0.4941}$

$= 0.70$

The F1-score of 70% indicates that the model achieved a balanced performance between precision and recall.

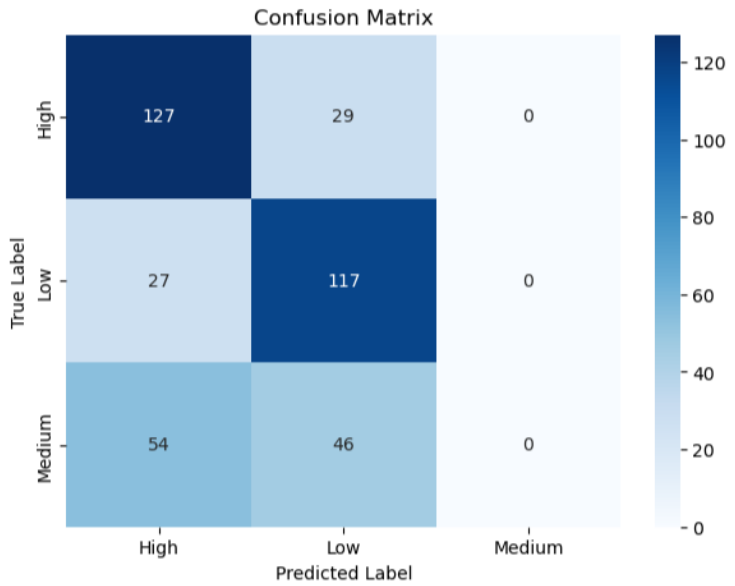


Fig 25: Confusion Matrix

The confusion matrix illustrates how the Random Forest model classified patients into Low, Moderate, and High relapse risk categories by comparing the actual patient risk levels with the predicted results generated by the system. The model achieved an overall accuracy score of 61%, meaning that 61 out of every 100 patient predictions were correctly classified. Most correct predictions appeared along the diagonal of the matrix, showing that the model successfully learned patterns associated with relapse risk. A few misclassifications occurred mainly within the Moderate-risk category, indicating that some patient behavioural patterns shared similarities across risk levels. Despite these minor errors, the confusion matrix demonstrates that the model was able to perform reasonably well in identifying patient relapse conditions.

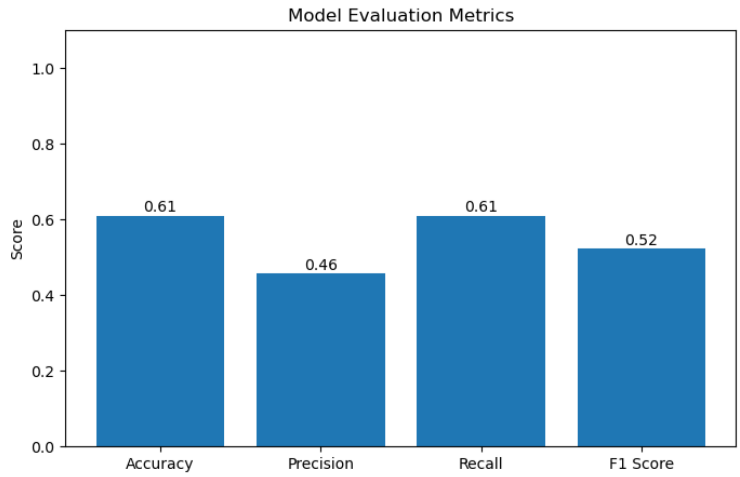


Fig 26 : Bar Chart of Model Evaluation Metrics of Accuracy, Precision, Recall and F1 Score

This graph illustrates the measures of performance of the Random Forest Relapse Prediction model using specific evaluation metrics. The accuracy score is 61%, implying that most patient relapse predictions by the model were accurate. The precision score of 46% means that whenever the model classified relapse risk categories, it was 46% of the time correct. The recall rate is 61%, meaning that the model was able to classify 61% of relapse risk categories in the dataset accurately. With an F1-score of 52%, the model has a balanced precision and recall. From the graph above, it can be concluded that though the model is not yet perfect, it can classify relapse risk patterns from patients' behavior data.

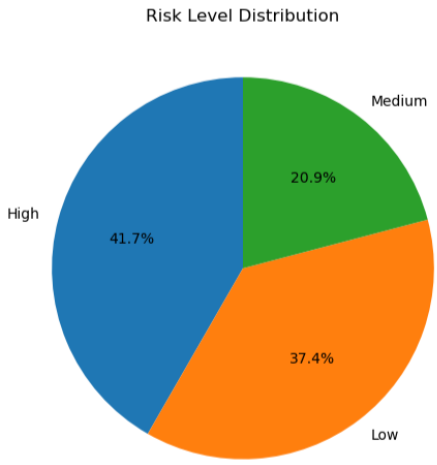


Fig 27: Risk Level Distribution Pie Chart

The pie chart illustrates the distribution of patients across the Low, Moderate, and High relapse risk categories used during model training and testing. The visualization shows that the dataset contained

patients from all relapse risk levels, allowing the model to learn different behavioural characteristics associated with recovery and relapse conditions. The varying proportions of each category helped the system identify relapse trends across multiple patient groups and contributed to the model's ability to make generalized predictions.

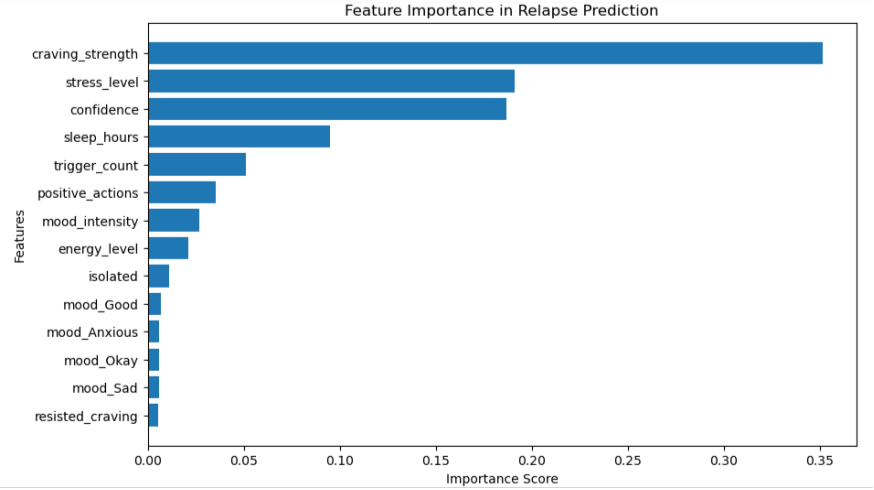


Fig 28: Feature Importance Graph

The feature importance graph presents the variables that contributed most significantly to predicting patient relapse risk within the Random Forest model. Features with higher importance values had a stronger influence on the model's prediction decisions because they contained behavioural patterns closely associated with relapse conditions. Variables such as stress level, cravings, mood level, sleep quality, medication adherence, and social interaction were selected because rehabilitation studies have shown that these factors strongly influence patient recovery progress and relapse likelihood. For example, high stress levels and strong cravings are commonly associated with increased relapse risk, while good sleep quality and proper medication adherence may reduce relapse chances. Features with larger importance scores therefore played a greater role in helping the model accurately classify patient relapse risk levels.

4.2.2 Evaluating the Machine Learning model used in Chatbot and Journal

The ML model, trained using Logistic Regression, achieved:

	precision	recall	f1-score	support
High	0.92	0.91	0.91	5075
Low	0.87	0.93	0.90	3633
Moderate	0.85	0.70	0.76	1083
accuracy			0.89	9791
macro avg	0.88	0.85	0.86	9791
weighted avg	0.89	0.89	0.89	9791

Fig 29: Results

1. Accuracy Calculation

Accuracy measures the proportion of total predictions that were correctly classified by the model.

$$Precision = \frac{4669 + 3161 + 921}{9791}$$

$$Precision = \frac{8751}{9791}$$

$$Precision = 0.89 = 89\%$$

2. Precision Calculation

Precision measures how many predicted classifications were actually correct.

$$Precision = \frac{4669 + 3161 + 921}{9791}$$

$$Precision = \frac{8751}{9791}$$

$$Precision = 0.89 = 89\%$$

3. Recall Calculation

Recall measures how many actual cases were correctly identified by the model.

$$Recall = \frac{4618 + 3379 + 758}{9791}$$

$$Recall = \frac{8755}{9791}$$

$$Recall = 0.89 = 89\%$$

4. F1-Score Calculation

The F1-score provides a balance between precision and recall.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$F1 = 2 \times \frac{0.89 \times 0.89}{0.89 + 0.89}$$

$$F1 = 0.89 = 89\%$$

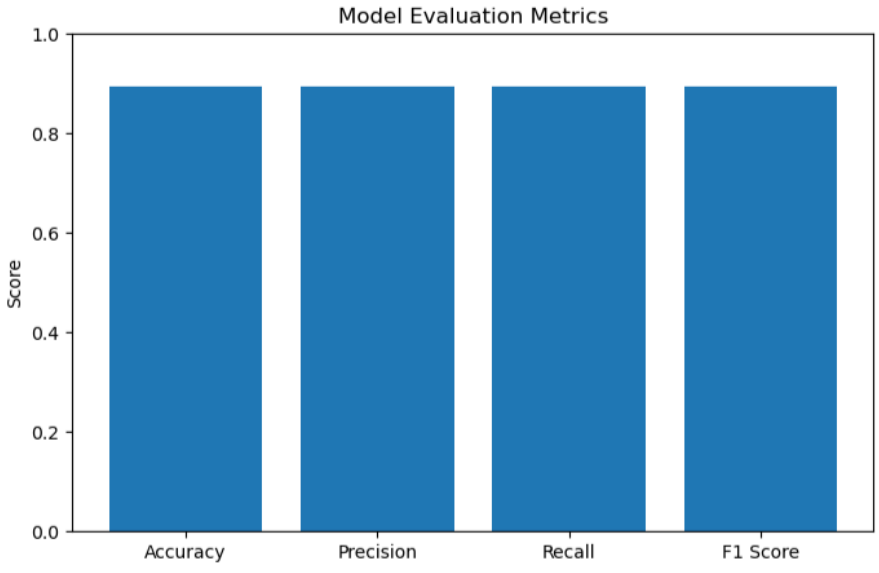


Figure 30: Machine Learning Prediction Metrics

The evaluation metrics chart shows that the model achieved strong performance across accuracy, precision, recall, and F1-score. This demonstrates that the classifier was reliable in identifying distress-related patient messages with high overall consistency.

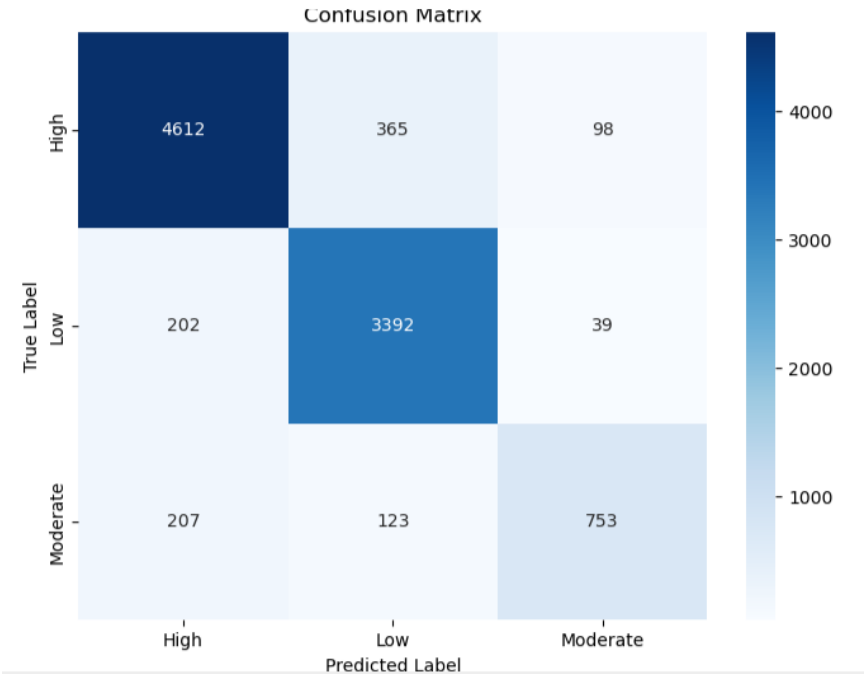


Figure 31: Confusion Matrics

The confusion matrix illustrates how well the model classified the different risk categories by comparing actual and predicted labels. Most messages were correctly classified, indicating that the model effectively learned patterns associated with low, moderate, and high emotional risk.

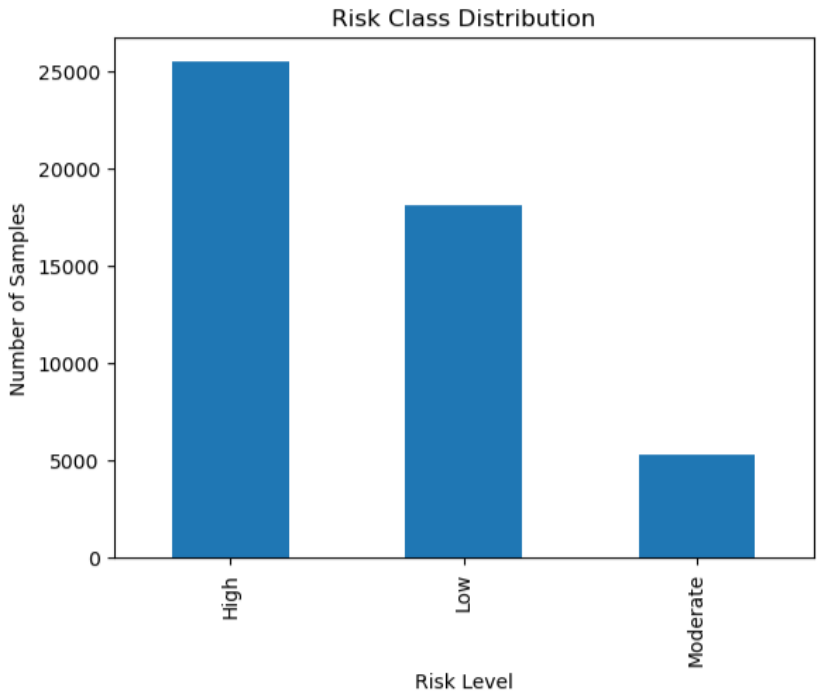


Fig 32: Risk Class Distribution Bar Chat

The risk distribution bar chart provides a visual comparison of the number of messages in each risk category. The results indicate that the dataset had a higher concentration of high-risk and low-risk messages compared to moderate-risk messages.

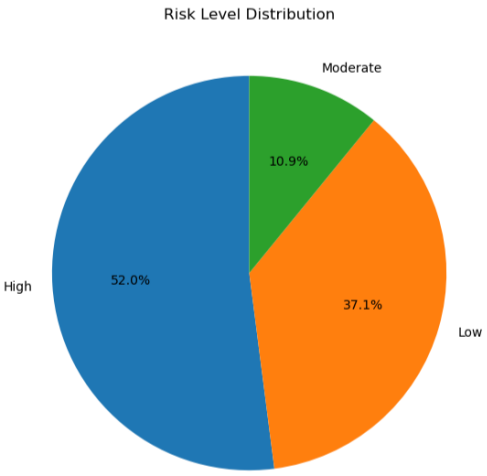


Fig 33: Risk Distribution Pie Chart

The risk distribution pie chart illustrates the proportion of low, moderate, and high-risk messages within the dataset. The chart shows that some risk categories contained more samples than others, which may have influenced the model's prediction performance.

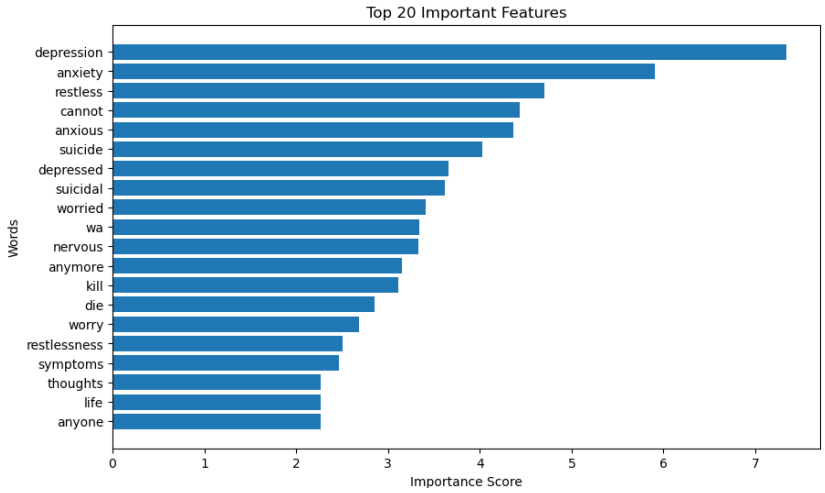


Fig 34: Feature Importance Diagram

The feature importance diagram highlights the words and textual features that contributed most significantly to predicting emotional distress and relapse risk within the chatbot and journal system. Features with higher importance values had a stronger influence on the model's prediction decisions because they were closely associated with emotional instability and relapse behaviour. Important words related to sadness, stress, anxiety, loneliness, hopelessness, cravings, and emotional distress were highly influential because they frequently appeared in high-risk patient messages. These features were selected because patient language and emotional expressions provide important indicators of psychological wellbeing and relapse risk. The model analysed the occurrence and significance of these words to determine patient emotional states, allowing the system to classify distress levels more accurately.

4.3 Analysis and interpretations of results

This section presents the analysis and interpretation of the two machine learning models used in the Drug and Substance Abuse Patient Relapse Prediction System. The first model was integrated into the chatbot and journal module to detect emotional distress levels from patient conversations and journal entries. The second model was integrated into the patient check-in form to predict relapse risk levels based on patient behavioral and monitoring data.

4.3.1 Chatbot and Journal Patient Relapse Prediction Model

The Chatbot and Journal Patient Relapse Prediction Model was evaluated using accuracy, precision, recall, and F1-score metrics, and the model achieved an overall score of 89% across all four evaluation measures, indicating strong and reliable performance. The accuracy score of 89% shows that the model correctly classified emotional distress and relapse risk levels in 89 out of every 100 predictions made, demonstrating its effectiveness in analyzing patient conversations and journal entries. The precision score of 89% indicates that when the model predicted a patient's emotional risk level, it was correct most of the time, meaning very few false positive classifications were produced. The recall score of 89% further demonstrates that the model successfully identified the majority of actual emotional risk cases, which is important in healthcare systems because it reduces the chances of missing patients who may require emotional support, counselling, or intervention. In addition, the F1-score of 89% reflects a strong balance between precision and recall, confirming that the model achieved both high prediction accuracy and consistent performance. The classification report also showed that the model performed well across all emotional risk categories, including High, Low, and Moderate classes, with only slight variations in prediction performance. Overall, the results indicate that the chatbot and journal model was highly effective in detecting emotional distress and behavioral relapse indicators, making it suitable for supporting healthcare personnel in continuous patient monitoring, early relapse detection, and timely patient intervention.

4.3.2 Check-In Form Relapse Prediction Model

The Check-In Form Relapse Prediction Model was evaluated using accuracy, precision, recall, and F1-score metrics to determine its effectiveness in predicting patient relapse risk levels from check-in form data. The model achieved an accuracy of 61%, precision of 46%, recall of 61%, and an F1-score of 52%, indicating moderate overall performance. The accuracy score shows that the model correctly predicted relapse risk levels in 61 out of every 100 cases, demonstrating its ability to identify behavioral patterns associated with patient relapse. The precision score of 46% indicates that some false positive classifications

were generated, meaning certain patients were incorrectly assigned to risk categories. However, in healthcare systems, identifying potentially risky patients is often prioritized to ensure early intervention and support. The recall score of 61% further demonstrates that the model successfully identified a reasonable number of actual relapse risk cases, reducing the chances of overlooking vulnerable patients who may require monitoring or counselling. The F1-score of 52% reflects a moderate balance between precision and recall, suggesting that while the model was capable of detecting relapse cases, there is still room for improvement in prediction consistency and classification accuracy. The classification report also showed that the model performed better in identifying High-risk and Low-risk patients compared to the Medium-risk category, which recorded lower prediction performance. Overall, the model demonstrated moderate predictive capability and provided useful support for patient monitoring and early relapse detection, although future improvements in data collection and model optimization could further enhance its reliability and performance.

4.4 A Summary of Research Findings

This section critically analyses the findings presented in this chapter, interpreting their significance in relation to existing studies and theories. It explains how the developed Drug and Substance Abuse Patient Relapse Prediction system confirms, extends, or differs from previous research in healthcare management systems, machine learning, and artificial intelligence in mental health support. Furthermore, the discussion addresses the effectiveness of the developed models, the chatbot and journaling functionalities, and the overall contribution of the system toward relapse prediction and patient monitoring.

4.4.1 AI-Powered Relapse Prediction and Early Risk Detection

The primary objective of this research was to develop an AI model for predicting relapse risk amoung drug and substance abuse patients. The findings confirm that the developed prediction models were able to classify patient risk levels based on behavioural, emotional, and psychological indicators collected through daily check-in forms, chatbot and journal entries. The system successfully identified different levels of patient risk, enabling early intervention and continuous patient monitoring.

The findings support existing studies which suggest that machine learning techniques can improve early detection of mental health conditions and substance abuse relapse risks through predictive analytics (Topol, 2019; Shatte et al., 2019). The successful implementation of predictive models within the system

demonstrates the practical application of artificial intelligence in healthcare environments, particularly in supporting healthcare professionals with data-driven decision-making processes. By analysing user responses and behavioural patterns, the system was capable of identifying warning signs associated with anxiety, depression, emotional distress, and potential relapse conditions.

The use of machine learning models for relapse prediction aligns with the principles of predictive analytics in healthcare, where historical and behavioural data are used to identify future health risks before severe outcomes occur. Similar studies by Esteva et al. (2019) and Rajpurkar et al. (2022) emphasize the growing role of AI systems in enhancing healthcare diagnosis, monitoring, and patient support services. This study extends these findings by applying AI techniques specifically to drug and substance abuse management within a patient-centered healthcare system.

The implementation of automated risk detection mechanisms significantly improves the ability of healthcare providers to monitor patients continuously without relying solely on manual observation. Traditional rehabilitation monitoring methods often depend heavily on periodic physical assessments and self-reporting, which may delay intervention. The developed system introduces a more proactive monitoring approach capable of providing real-time feedback and identifying high-risk patients earlier. This contributes toward improving patient support and potentially reducing relapse occurrences.

4.4.2 System Usability, User Acceptance, and Reliability

System testing was conducted to evaluate the usability, functionality, and reliability of the developed AI-powered patient management system. The findings indicate that the system operated successfully under controlled testing conditions and that the interface was user-friendly and accessible. The integration of journaling functionality, chatbot interaction, and relapse prediction tools created a centralized platform capable of supporting both patients and healthcare administrators.

The developed system demonstrated practical usefulness by assisting in patient monitoring, emotional analysis, and relapse prediction while maintaining a simple and interactive interface for users.

4.4.3 Study limitations

The study has outlined some limitations that played a role in shaping the interpretation, assessment, and validity of the findings as pointed out in the limitations presented in Chapter 1 and gaps in the literature presented in Chapter 2.

4.4.3.1 Limited Financial Resources for Data Collection:

Limited Financial Resources had a considerable impact on the process of collecting data for this study since it restricted the researcher's capacity to collect large quantities of primary patient data in actual rehabilitation facilities for use in the training and testing of the machine learning models. For this reason, part of the datasets used in this study were simulated other datasets were collected through secondary sources like Kaggle, which were useful for the development of the chatbot and journal analysis functionalities.

Chapter Summary

This chapter presented the analysis and interpretation of results obtained from the evaluation of the Drug and Substance Abuse Patient Relapse Prediction System and its machine learning models. The findings showed that the Chatbot and Journal Patient Relapse Prediction Model achieved strong and reliable performance, while the Check-In Form Relapse Prediction Model demonstrated moderate predictive capability in identifying patient relapse risks. The results indicate that the system can effectively support healthcare personnel in patient monitoring, emotional risk detection, and early relapse intervention. Overall, the study demonstrated the potential of integrating artificial intelligence into substance abuse patient management to improve monitoring and decision-making processes.

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter highlights the major findings of the study and the assessment on the success of achieving the goals of this study. Improvements that should be made for future studies are also recommended based on the results of this study.

Objective 2: To develop a Random Forest and NLP-based machine learning algorithm to predict relapse of DSA patients at FASAA.

5.2 Major conclusions drawn

It was established that the conventional approaches used in monitoring of recovering drug and substance abuse patients are highly manual and inconsistent hence lack early detection of relapses. Interviews and observations conducted at FASAA indicated that some patients struggle to openly communicate their emotional and psychological challenges during face-to-face counselling sessions with social workers. In some cases, fear of stigma, shame and lack of confidence prevent patients from fully expressing the underlying problems contributing to their drug and substance abuse behaviors. As a result, important information needed for effective treatment and recovery support may remain hidden.

The development of the Drug and Substance Abuse Patient Relapse Prediction System demonstrated that machine learning technologies can assist healthcare professionals by identifying high-risk patients before relapse occurs. The Random Forest model successfully analyzed structured patient check-in data such as mood changes, stress levels, cravings and behavioral patterns to classify patients according to relapse risk levels. At the same time, the NLP-based model effectively analyzed patient journal entries and chatbot conversations to identify emotional distress, negative thoughts and possible relapse indicators from patient-generated text.

The introduction of the journal and chatbot features created a safe and less intimidating environment where patients could express their feelings, emotions and personal struggles more freely. This allowed the system to capture important hidden information that patients may not disclose directly to social workers. The system further enhanced intervention by displaying relapse alerts together with the possible contributing factors causing the high relapse risk. This enabled social workers and caregivers to better understand the patient's condition and provide more targeted emotional and psychological support.

Based on these findings, several recommendations are made.

5.3 Recommendations

i. FASAA should fully implement the Drug and Substance Abuse Patient Relapse Prediction System to improve early intervention and patient relapse prediction.
ii. Healthcare professionals and social workers should be trained on how to interpret AI-generated relapse predictions and alerts to support informed decision-making.
iii. Patients should be educated and guided on how to use the system effectively, particularly those with limited digital literacy or low educational backgrounds, to ensure that all patients can benefit from the platform regardless of their technical skills.
iv. More real-world patient data should be collected to further improve the performance and generalization of both the Random Forest and NLP models.
v. Proper cybersecurity and data privacy measures must be put in place to ensure confidentiality and security of patient data.

5.4 Future Works

While the Drug and Substance Abuse Patient Relapse Prediction System has achieved remarkable milestones, there are areas for ongoing development:

i. Deep learning algorithms should be explored to enhance predictive accuracy of the system.
ii. Future studies may also explore multilingual chatbot support, real-time mobile notifications and integration with electronic health record systems to improve accessibility and scalability.

Summary

This research was able to prove that the application of Random Forest and machine learning models using Natural Language Processing in drug and substance abuse patients can contribute greatly to the accuracy in predicting relapses among patients, and consequently, in improving the monitoring and intervention processes. This research shows the potential usefulness of applying such AI models in health service provision.

References

Acion, L., Kelmansky, D., Laan, M.D.V., Sahker, E., Jones, D.S., Arndt, S., 2017. Use of a machine learning framework to predict substance use disorder treatment success. PLoS ONE 12. https://doi.org/10.1371/journal.pone.0175383

Adam, M.T.P., Gregor, S., Hevner, A., Morana, S., 2021. Design Science Research Modes in Human-Computer Interaction Projects. AIS Transactions on Human-Computer Interaction 13, 1–11. https://doi.org/10.17705/1thci.00139

Balise, R.R., Grealis, K., Brandt, L., Luo, S.X., Odom, G., Jean-Baptiste, G., Feaster, D.J., 2026. Applying machine learning in predicting medication treatment outcomes for opioid use disorder. Journal of Substance Use and Addiction Treatment 182, 209847. https://doi.org/10.1016/j.josat.2025.209847

Curtis, B., Giorgi, S., Ungar, L., Vu, H., Yaden, D., Liu, T., Yadeta, K., Schwartz, H.A., 2023. AI-based analysis of social media language predicts addiction treatment dropout at 90 days. Neuropsychopharmacol. 48, 1579–1585. https://doi.org/10.1038/s41386-023-01585-5

Efthimiou, O., Seo, M., Chalkou, K., Debray, T., Egger, M., Salanti, G., 2024. Developing clinical prediction models: a step-by-step guide. BMJ. https://doi.org/10.1136/bmj-2023-078276

Faltushanskiy, Z., Herrold, A.A., Werby, J., Betteridge, E.M., Angres, D., 2024. Personality predicting relapse: A facet analysis of the NEO PI-R. American Journal on Addictions 33, 58–64. https://doi.org/10.1111/ajad.13477

Farooq, N., Riaz, Dr.S., 2022. Psychosocial Factors as the Determinants of Relapse in Individuals with Substance Use Disorder. IJIST 4, 97–104. https://doi.org/10.33411/IJIST/2022040611

Ghitza, U., Ghitza, U., Sparenborg, S., Tai, B., 2011. Improving drug abuse treatment delivery through adoption of harmonized electronic health record systems. SAR 125. https://doi.org/10.2147/SAR.S23030

Gowin, J.L., Ball, T.M., Wittmann, M., Tapert, S.F., Paulus, M.P., 2015. Individualized relapse prediction: Personality measures and striatal and insular activity during reward-processing robustly predict relapse. Drug and Alcohol Dependence 152, 93–101. https://doi.org/10.1016/j.drugalcdep.2015.04.018

Kabisa, E., Biracyaza, E., Habagusenga, J. d'Amour, Umubyeyi, A., 2021. Determinants and prevalence of relapse among patients with substance use disorders: case of Icyizere Psychotherapeutic Centre. Subst Abuse Treat Prev Policy 16, 13. https://doi.org/10.1186/s13011-021-00347-0

Lauvsnes, A.D.F., Gråwe, R.W., Langaas, M., 2022. Predicting Relapse in Substance Use: Prospective Modeling Based on Intensive Longitudinal Data on Mental Health, Cognition, and Craving. Brain Sciences 12, 957. https://doi.org/10.3390/brainsci12070957

Marlatt, G.A., Witkiewitz, K., Donovan, D.M., 2005. RELAPSE PREVENTION Relapse Prevention for Alcohol and Drug Problems Relapse Prevention for Alcohol and Drug Problems.

Marsch, L.A., 2020. Digital health and addiction. Current Opinion in Systems Biology 20, 1–7. https://doi.org/10.1016/j.coisb.2020.07.004

Matutu, V., Mususa, D., 2019. Drug and Alcohol Abuse Among Young People in Zimbabwe: A Crisis of Morality or Public Health Problem. SSRN Journal. https://doi.org/10.2139/ssrn.3489954

Mortazavi, L., MacNiven, K.H., Knutson, B., 2024. Blunted Neurobehavioral Loss Anticipation Predicts Relapse to Stimulant Drug Use. Biological Psychiatry 95, 256–265. https://doi.org/10.1016/j.biopsych.2023.07.020

Nawi, A.M., Ismail, R., Ibrahim, F., Hassan, M.R., Manaf, M.R.A., Amit, N., Ibrahim, N., Shafurdin, N.S., 2021. Risk and protective factors of drug abuse among adolescents: a systematic review. BMC Public Health 21, 2088. https://doi.org/10.1186/s12889-021-11906-2

Odejide, A.O., 2006. Status of Drug Use/Abuse in Africa: A Review. Int J Ment Health Addiction 4, 87–102. https://doi.org/10.1007/s11469-006-9015-y

Onaolapo, O.J., Olofinnade, A.T., Ojo, F.O., Adeleye, O., Falade, J., Onaolapo, A.Y., 2022. Substance use and substance use disorders in Africa: An epidemiological approach to the review of existing literature. World Journal of Psychiatry 12, 1268–1286. https://doi.org/10.5498/wjp.v12.i10.1268

Peffers, K., Tuunanen, T., Gengler, C.E., Rossi, M., Hui, W., Virtanen, V., Bragge, J., 2006. THE DESIGN SCIENCE RESEARCH PROCESS: A MODEL FOR PRODUCING AND PRESENTING INFORMATION SYSTEMS RESEARCH.

Ranganathan, P., Caduff, C., 2023. Designing and validating a research questionnaire - Part 1. Perspectives in Clinical Research 14, 152–155. https://doi.org/10.4103/picr.picr_140_23

Spencer, M., Garnett, M., Miniño, A., 2023. Drug Overdose Deaths in the United States, 2002-2022. National Center for Health Statistics (U.S.). https://doi.org/10.15620/cdc:135849

Tai, B., 2012. Electronic health records: essential tools in integrating substance abuse treatment with primary care. SAR 1. https://doi.org/10.2147/SAR.S22575

The University of Newcastle, Adam, Marc.T.P., Gregor, S., Australian National University, Hevner, A., University of South Florida, Morana, S., Saarland University, 2021. Design Science Research Modes in Human-Computer Interaction Projects. THCI 1–11. https://doi.org/10.17705/1thci.00139

Theofanidis, D., Fountouki, A., 2019. LIMITATIONS AND DELIMITATIONS IN THE RESEARCH PROCESS. https://doi.org/10.5281/ZENODO.2552022

Vafaie, N., Kober, H., 2022. Association of Drug Cues and Craving With Drug Use and Relapse: A Systematic Review and Meta-analysis. JAMA Psychiatry 79, 641. https://doi.org/10.1001/jamapsychiatry.2022.1240

Willie, M.M., 2024. Population and Target Population in Research Methodology. Golden Ratio of Social Science and Education 4, 75–79. https://doi.org/10.52970/grsse.v4i1.405

Wu, L.-T., Payne, E.H., Roseman, K., Kingsbury, C., Case, A., Nelson, C., Lindblad, R., 2019. Clinical Workflow and Substance Use Screening, Brief Intervention, and Referral to Treatment Data in the Electronic Health Records: A National Drug Abuse Treatment Clinical Trials Network Study. eGEMs 7, 35. https://doi.org/10.5334/egems.293

Wyant, K., Sant'Ana, S.J., Fronk, G.E., Curtin, J.J., 2024. Machine learning models for temporally precise lapse prediction in alcohol use disorder. Journal of Psychopathology and Clinical Science 133, 527–540. https://doi.org/10.1037/abn0000901

Xu, J., Li, X., Liang, X., Zhang, X., Ying, G., Liu, J., 2025. The risk factors for relapse behavior in individuals with substance use disorders: An interpretable machine learning study. Journal of Affective Disorders 387, 119520. https://doi.org/10.1016/j.jad.2025.119520

Yamashita, A., Yoshioka, S. ichi, Yajima, Y., 2021. Resilience and related factors as predictors of relapse risk in patients with substance use disorder: a cross-sectional study. Substance Abuse: Treatment, Prevention, and Policy 16. https://doi.org/10.1186/s13011-021-00377-8

Yuan, W., Chen, M., Wang, D.W., Li, Q.H., Yin, Y.Y., Li, B., Wang, H.R., Hu, J., Gong, Y.D., Yuan, T.F., Yu, T.G., 2024. Computational markers of risky decision-making predict for relapse to alcohol. European Archives of Psychiatry and Clinical Neuroscience 274, 353–362. https://doi.org/10.1007/s00406-023-01602-0