

Identification of Opioid Patients by Employing Structured and Unstructured Data from MIMIC-III Database

Saddam Al Amin
Student Id: 012201022

A Thesis
in
The Department
of
Computer Science and Engineering



Presented in Partial Fulfillment of the Requirements
For the Degree of Master of Science in Computer Science and Engineering
United International University
Dhaka, Bangladesh
May 2023
© Saddam Al Amin, 2023

Approval Certificate

This thesis titled " **Identification of Opioid Patients by Employing Structured and Unstructured Data from MIMIC-III Database**" submitted by [**Saddam Al Amin**], Student ID: [**012201022**], has been accepted as Satisfactory in fulfillment of the requirement for the degree of Master of Science in Computer Science and Engineering on.

Board of Examiners

1.

Supervisor

Dr. Md. Saddam Hossain Mukta
Associate Professor & Undergraduate Coordinator
Dept. of Computer Science and Engineering
United International University
Dhaka, Bangladesh

2.

Head Examiner

Dr. Swakkhar Shatabda
Professor & Director - IQAC
Dept. of Computer Science and Engineering
United International University
Dhaka, Bangladesh

3.

Examiner-I

Dr. Khondaker Abdullah Al Mamun
Professor & Director - AMIS Labs
Dept. of Computer Science and Engineering
United International University
Dhaka, Bangladesh

4.

Examiner-II

Dr. Dewan Md. Farid
Professor
Dept. of Computer Science and Engineering
United International University
Dhaka, Bangladesh

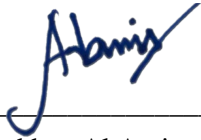
5.

Ex-Officio

Dr. Mohammad Nurul Huda
Professor & Director - MSCSE Program
Department of Computer Science and Engineering
United International University
Dhaka, Bangladesh

Declaration

This is to certify that the work entitled “**Identification of Opioid Patients by Employing Structured and Unstructured Data from MIMIC-III Database**” is the outcome of the research carried out by me under the supervision of Dr. Md. Saddam Hossain Mukta, Associate Professor, Dept. of CSE, UIU.



Saddam Al Amin

Student ID: 012201022

MSCSE Program

Department of Computer Science and Engineering

United International University

Dhaka, Bangladesh

In my capacity as supervisor of the candidate’s thesis, I certify that the above statements are true to the best of my knowledge.

Dr. Md. Saddam Hossain Mukta

Associate Professor & Undergraduate Coordinator

Dept. of Computer Science and Engineering

United International University

Dhaka, Bangladesh

Abstract

The opioid crisis has led to an increased number of drug overdoses in recent years. Several approaches have been established to predict opioid prescription by health practitioners. However, due to the complex nature of the problem, the accuracy of such methods is not yet satisfactory. Dependable and reliable classification of opioid dependent patients from well-grounded data sources is essential. Majority of the previous studies do not focus on the users' mental health association for opioid intake classification. These studies do not also employ the latest deep learning based techniques such as attention and knowledge distillation mechanism to find better insights. This thesis investigates the opioid classification problem by using machine learning and deep learning based techniques. We used structured and unstructured data from the MIMIC-III database to identify intentional and unintentional intake of opioid drugs. We selected 455 patient instances and used traditional machine learning and deep learning to predict intentional and accidental users. We obtained 95% and 64% test accuracy to predict the intentional and accidental users from the structured and unstructured datasets, respectively. We also achieve a distilled knowledge based test accuracy of 76.44% from the integrated above two models. Our research includes an ablation analysis and new insights related to opioid patients are extracted.

Keywords: Opioid intake, mental illness, MIMIC-III database, machine learning, deep learning

Acknowledgement

I want to express my heartfelt gratitude to my thesis supervisor Dr. Saddam Hossain Mukta, who gave me the opportunity to do this incredible research on classifying opioid patients. At the same time, I express my special thanks to the Computer Science department of United International University, Dhaka, Bangladesh, for their support and cooperation.

Publication List

The main contributions of this research is published in IEEE Access journal which has an impact factor of 3.47.

Journal Articles

1. Al Amin, S., Mukta, M. S. H., Saikat, M. S. M., Hossain, M. I., Islam, M. A., Ahmed, M., & Azam, S. (2022). Data Driven Classification of Opioid Patients Using Machine Learning–An Investigation. IEEE Access, 11, 396-409.

<https://www.hackerrank.com/domains>

Table of Contents

Table of Contents	viii
List of Figures	ix
List of Tables	x
1 Introduction	1
1.1 Project Overview	1
1.2 Methodology	2
1.3 Project objectives	3
1.4 Organization of the Report	3
2 Background	4
2.1 Preliminaries	4
2.2 Literature Review	8
2.3 Gap Analysis	10
3 Methodology	11
3.1 Data Description and Statistics	11
3.2 Dataset Preparation	12
3.2.1 Pre-processing	12
3.2.2 ICD-9 codes	12
3.2.3 Ground truth	14
3.2.4 Agreement Analysis among Annotators	15
3.2.5 Case Identification Based on Keywords	16
3.3 Model Building	17
4 Model building with Ablation Studies	18
4.0.1 Model Building	18
4.0.2 Combining models by Knowledge Distillation	21
4.1 Results	23
4.1.1 From structured dataset	23
4.1.2 From unstructured data	24
4.2 Ablation study	24
4.2.1 Ablation study 1: changing hidden layers	24
4.2.2 Ablation study 2: changing batch size	25
4.2.3 Ablation study 3: changing optimizer	25
4.2.4 Ablation study 4: changing learning rate	26
4.2.5 Ablation study 5: changing dropouts	26
4.2.6 Summary of the ablation study	27
4.2.7 Selecting the best model	28
5 Discussion	29

6 Conclusion 32

6.1 Limitation 32

6.2 Future Work 32

References 38

List of Figures

2.1	Transformer architecture	5
2.2	LSTM architecture	6
2.3	BiLSTM architecture	7
2.4	CNN architecture	7
3.1	Methodology of opioid misuse prediction from	11
3.2	Missing Value handling technique selection	13
3.3	Keyword base steps of Predicting Intentional/Unintentional Opioid Cases	14
4.1	Combining models (structured+unstructured) by knowledge distillation.	23
5.1	Frequently found problems, test and treatments name for unintentionally opioid users	30
5.2	Frequently Found problems, test and treatments name for intentionally opioid users	30

List of Tables

1.1	Number of Opioid-related deaths in the last decade (2010-2020)	2
2.1	Comparative Analysis Table on Related Opioid Research	9
3.1	Statistics of the MIMIC-III database.	12
3.2	Association with Behavioral Health and Opioid Intake	13
3.3	Statistics of our filtered dataset.	15
4.1	Performance of different classifiers to predict opioid behavior by using structured dataset. . . .	18
4.2	Summary of the models that have been used to classify unstructured dataset.	21
4.3	Performance of different classifiers to predict opioid behavior from unstructured eventnotes. . .	24
4.4	Performance measured by changing hidden layers.	25
4.5	Performance Measured by Changing Batch Size	26
4.6	Performance Measured by Changing Optimizer	26
4.7	A method to prevent neural networks from overfitting is dropout regularization on Performance measured by changing learning rates.	27
4.8	Performance Measured by Changing dropout percentage	27

Chapter 1

Introduction

Opioid analgesics are generally used to alleviate severe and chronic pain in patients. Doctors and other health care practitioners prescribe opioids in large numbers, especially in the United States of America (USA). According to the Centers for Disease Control and Prevention (CDC), the approximate cost of opioid abuse in the United States is \$78.5 billion per year [1]. The number of opioid prescriptions in the United States is very high; research found that around 153 million opioid drugs were prescribed in 2019 [2]. This problem is now a health epidemic and a concern among health practitioners, researchers, and policymakers. For instance, table 1.1 describes opioid-related deaths are increasing yearly despite many strategies to curb this uptick. Hence, much research is needed in this domain to create awareness and thorough understanding among various professions. In this research, we work with a well-established dataset to classify opioid patients.

1.1 Project Overview

Opioids are a class of drugs prescribed as painkillers, but they are heavily overused due to their addictive nature. Several studies [3,4] have described that patients get these medications not to control pain; but because they are dependent on them. This can also result in an overdose. In our study, we use machine learning techniques to predict users' opioid misuse patterns from both structured data (i.e., demographic information, gender, ethnicity, etc.) and unstructured data (i.e., chronological medical history and eventnotes).

Barkley et al. [5] found that intentional overdoses correlated with a depression. Other studies [6,7] found that the rate of intentional drug use among adolescents is worrying. Prince et al. [8] found that there is a direct connection between taking drugs and mental illness. Jones et al. [9] also found that opioid use disorder (OUD) is associated with mental disorders. There appears to be a direct relationship [10,11] between mental illness and drug abuse which needs further investigation. In the studies mentioned above, most authors conduct research on a specific aspect of the opioid problem, such as particular age groups or demographics [12–14]. The database we utilize is a good source of data which includes demographic, ethnicity, medical condition and age variables to study the problem. Previous studies did not use contextual analysis based on natural language processing (NLP) techniques of the patients' event notes, and medical history.

Deep learning and Machine Learning have gained popularity in the healthcare applications [15–17]. However, the current opioid risk assessment tools are insufficient in terms of predictability and automatic contextual

Table 1.1: Number of Opioid-related deaths in the last decade (2010-2020)

Year	Number of Deaths
2010	38329
2011	41340
2012	41502
2013	43982
2014	47055
2015	52404
2016	63632
2017	70237
2018	67367
2019	70630
2020	81000

analysis based on patients' historical data ¹ . Furthermore, clinicians should be offered tools that allow determination of patients' risk of misuse before administering opioids. Considering that opioid misuse is a medical problem impacting people's health and economy, investigating the problem based on a Machine learning approach can be useful. The database that we work with has data which could be utilized to identify opioid patients. In the light of the above discussion, previous studies find an association between mental health and opioid intake. In some other studies, researchers consider demographics (e.g., age, ethnicity, etc.) for finding opioid associations. Therefore it is important to utilize the above features as the predictors of opioid intaking early warning systems. In addition to this, users' historical data provides a contextual cue for users' future behavior. Previous studies rarely employ the latest deep learning based NLP techniques such as attention and knowledge distillation mechanism from the contextual signals which can unveil better insight for the researchers.

1.2 Methodology

In this thesis, we use data from the MIMIC-III database [18], from which we have identified the opioid cases based on keyword identification. We identify relevant tables (i.e., schemas) from the database and select 41 features which are relevant to our study. Based on the keywords and patients' history, we identify which patients take opioids intentionally. In this way, we label our dataset as opioid intake 'YES'/'NO'. Later, we build a structured (i.e., tabular) dataset. To strengthen the model, we also incorporate an unstructured dataset. As training an unstructured dataset is complex and challenging, we apply deep learning based NLP techniques. For each patient, we analyze their historical data (i.e., event notes/unstructured data), and we convert the data using word embedding and attention based LSTM techniques. Since our patients data is already labelled, we train the unstructured data with the deep learning based technique mentioned above. In this study, we obtain a higher performance model by using the structured dataset while the model using unstructured dataset shows weaker

¹<https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates>

results. To build a combined model, we apply knowledge distillation technique where structured dataset shows the higher capacity network and then, we transfer the knowledge to the weaker unstructured dataset.

Our study further investigates whether a pattern of opioid use has any connection with users' mental health statuses and other socio-economical determinants. Classification of opioid patients and their mental health is important, considering the number of overdose deaths per year and the financial consequences of opioid addiction [19]. Our study may benefit society in a number of ways, such as early detection of intentional and unintentional opioid misuse, reducing the effect of aggressive marketing by pharmaceutical companies which profit from pain medication use, and better surveillance of opioid misuse by authorities and stakeholders.

1.3 Project objectives

The main contributions of this study are:

1. We build a dataset by using the MIMIC-III database for predicting opioid misuse.
2. We investigate the relationship between mental health and opioid misuse by patients from their structured and unstructured data (patients' clinical event notes).
3. We develop traditional and deep learning based supervised models to predict intentional and unintentional Opioid users, using an attention based mechanism.

1.4 Organization of the Report

In Section IV, we discuss the related work, which includes preliminaries, literature review, gap analysis, and a summary table. Section V outlines the methodology containing data collection techniques to process of model building; Section VI emphasize again on the data collection in details, trailed by Section VII that provides the broad information on model building and performance of our model. In the VIII, we added the Ablation study. Lastly, in the IX section, we talked about challenges, the future of work, and ended the study with a conclusion.

Chapter 2

Background

In this chapter, first we present preliminaries to understand our study. Next, we describe the relevant works which were related to our research.

2.1 Preliminaries

Opioids: Opioids refer to a group of medications that are naturally present in the opium poppy plant that functions in a brain to cause a number of conditions, including pain reduction in certain cases. As a result, they are often used in medicine for pain control like anesthesia. By suppressing pain indicators in the body and the brain, prescription drugs are often used to relieve average to grave pain. Non-medically, (opioids) are often used for their euphoric properties. So, apart from pain relieving, opioids may also make certain patients feel relaxed, happy, or "high." Nevertheless, the addiction and withdrawal syndrome can be very dangerous . Constipation, confusion, nausea, slowed breathing, and drowsiness are some of the adverse consequences that may occur. Some commonly used opioids are: Methadone, Cocaine, Vicodin, Morphine, Fentanyl.

MIMIC-III: The MIMIC-III is a medical database containing a large amount of patient data, in particular, 40,000 records. It is a publicly available data having structured data with CSV format, and unstructured data in a table named 'NOTEEVENTS'. In total, there are 26 tables and we set up an SQL environment to run queries on thirteen schemas out of those twenty-six tables to aggregate necessary features in order to develop a structured opioid patient dataset.

Text Encoder and Decoder: The purpose of using a text encoder is to encode text. Text encoding is a process of converting meaningful text(s) into a vector representation by maintaining context and the relationship between words and sentences [20]. Another purpose of a text encoder is to generate fixed/same-size vector representation. This encoded vector(s) gets used as the input(s) of different algorithms/classifiers/models (e.g. LSTM, RNN, CNN, etc.) . Because of these reasons, a machine can understand the pattern associated with any text.

On the other hand, a decoder is a stack of several recurrent units where each predicts an output at a time step 't'. Each of the recurrent units accepts a hidden state from the previous unit and produces an output as its own hidden state.

Transformer: A transformer is a deep learning model that uses the self-attention process and weights the importance of each component of the input data differently. Liu et al., 2021 [21] say that it is used a lot in computer vision and natural language processing.

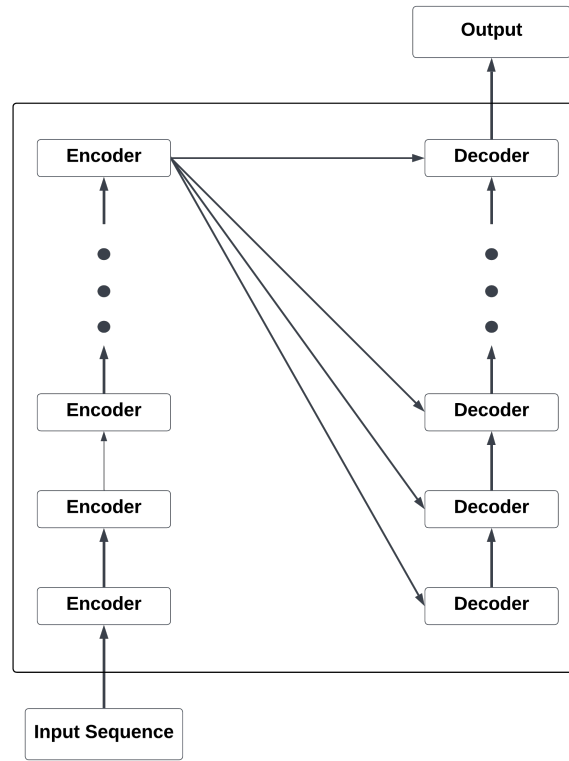


Figure 2.1: Transformer architecture

Transformers are used in activities like translation and text summarising to handle sequential input data, such as natural language. In contrast to RNNs, transformers process all of the input at once [22]

A Transformer is a novel type of neural architecture that uses the attention mechanism to encode the input data as potent characteristics. In essence, the visual transformers compute both representations and their relationships after splitting the input pictures into numerous local patches. The granularity of the patch division is insufficient for digging aspects of objects in various scales and locations since natural pictures are highly complex with plentiful detail and color information [23].

LSTM: In the field of data science, sequence prediction is regarded as among the most challenging issues to resolve. Long-term short-term memory networks (LSTMs) have been determined to be the most efficient solution for nearly all of these sequence prediction challenges. LSTM is an effective and scalable model for several learning problems related to sequential data.

In many aspects, LSTMs are superior to RNNs and traditional feed-forward neural networks. They can remember patterns that are important to them for long periods of time [24].

An artificial neural network called Long Short-Term Memory (LSTM) is employed in deep learning and artificial intelligence. Such a recurring neural network may analyze complete data sequences, including voice and videos, in addition to single data points (such as photos). The most frequently used neural network in the 20th century is LSTM. The impact of the LSTM network has been significant in language modeling, speech-to-text transcription, machine translation, and other applications [24].

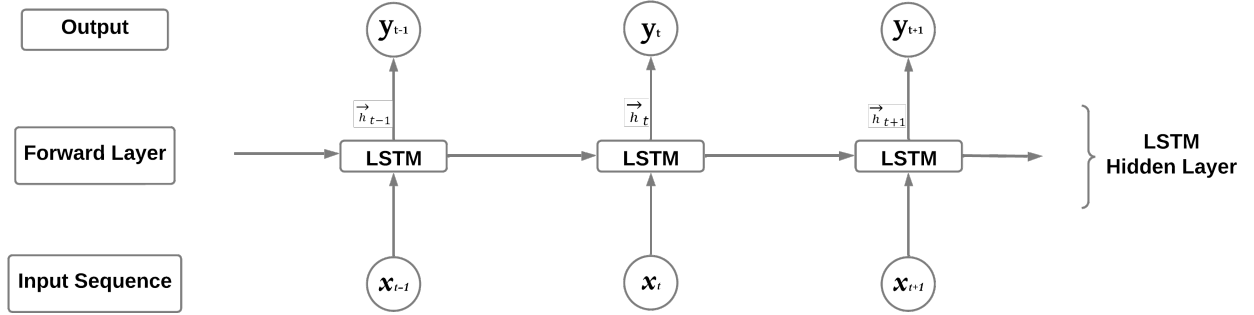


Figure 2.2: LSTM architecture

BiLSTM: A bi-directional LSTM, also known as a biLSTM, is a sequence processing model composed of two LSTMs, one of which takes the input forward and the other backward [25] found that using the LSTM twice makes it easier for the model to learn long-term correlations. This makes the model more accurate.

The embedding layer, which makes up the second layer of our suggested architecture, converts each index, which stands for a distinct word in the data set, into a real-valued feature vector. An embedding matrix is created by piling these real-valued vectors together. The BiLSTM output protocol is designed for the forward and backward states, which allows for a final summary of the BiLSTM output. The global maximum and global average pooling layers receive the outputs of the BiLSTM layer at the same time. Hameed and Garcia-Zapirain [26] say that the first and second layers take the maximum and average values of each feature in the BiLSTM layer.

Variants of Attention: Attention is a method that simulates cognitive attention in neural networks. The purpose of the impact is to encourage the network to pay greater attention to the minute but crucial portions of the input data by enhancing some portions while reducing others. Gradient descent is used to teach an algorithm how to figure out, based on the context, which part of the data is more important than another.

In the 1990s, processes resembling attention were first described under the titles multiplicative modules, sigma pi units, and hypernetworks. In contrast to normal weights, which must remain set throughout runtime, "soft weights" can alter during runtime, giving them flexibility. Attention is used for language processing in transformers, memory in neural Turing machines, reasoning in discrete neural computers, and multimodal data processing (audio, pictures, film, and writing) in sensory perceptions [27].

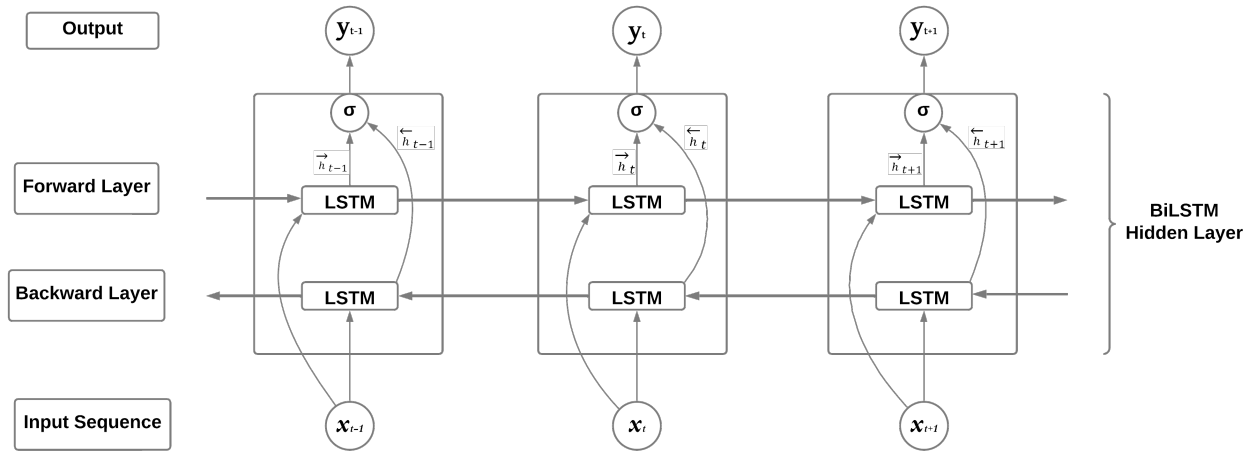


Figure 2.3: BiLSTM architecture

Convolutional Neural Networks (CNNs) have evolved over the past ten years into the mandated standard for many computer vision and machine learning activities. Convolutional and subsampling layers alternate in feed-forward artificial neural networks (ANNs) known as CNNs.

Abnormality detection and identification in power electronics, structural health monitoring, and electrical motor fault detection are just a few applications where 1D CNNs have recently been proposed and have already attained state-of-the-art performance levels. Due to the straightforward and compact architecture of 1D CNNs, which only execute 1D convolutions, a real-time and affordable hardware implementation is possible, which is a significant benefit [28].

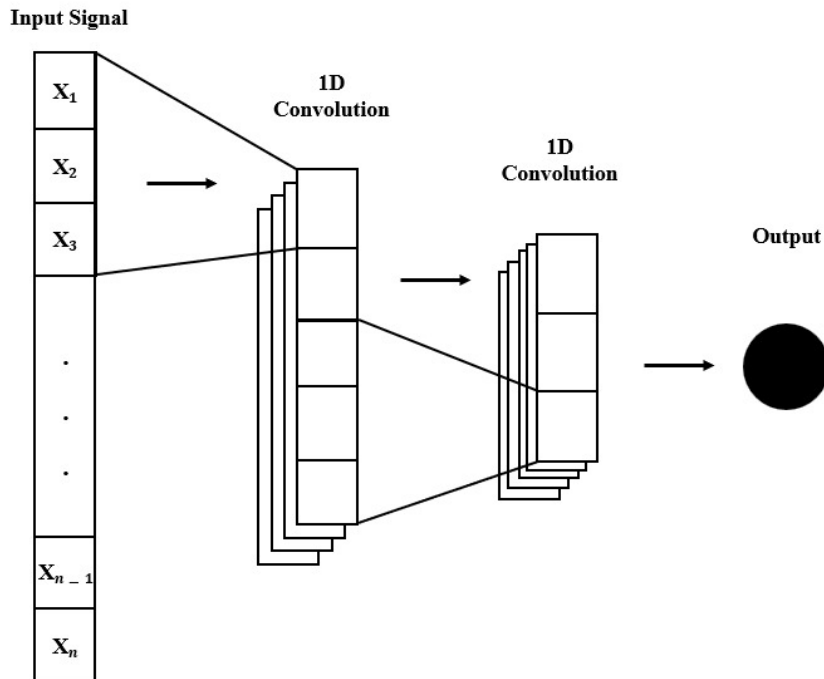


Figure 2.4: CNN architecture

CNN-1D: Recently, it has been shown that deep learning techniques, such as recurrent neural networks and one-dimensional convolutional neural networks, or CNNs, can do advanced activity recognition tasks with little or no data feature engineering and by learning features from raw data.

1 D CNN can determine an individual's activity from accelerometer data, such as whether they are standing, walking, leaping, etc. This data is in two dimensions. Time-steps make up the first dimension, while the acceleration values along three axes make up the second [29].

2.2 Literature Review

Prior research has shown that using opioids and benzodiazepines increases the risk of an overdose fatality compared to opioids use alone [30]. The authors described trends in intentional abuse of opioid analgesics, benzodiazepines, or both, from 2001 to 2014. They then calculated the increased risk of mortality associated with the abuse or misuse of the combination of opioid analgesics and benzodiazepines relative to opioid analgesic abuse or misuse alone. Barkley et al. [5] investigated the characteristics of the individuals who died from intentional drug overdoses compared to unintentional overdoses. They found that intentional overdoses are associated with depression. Another study [6] investigated the demographic characteristics related to intentional opioid usage. A study based on information provided by three poison control centers in 2002–2014 in Ohio [7] showed that the rate of intentional drug use among adolescents is alarming, and that there is a need for more research into the misuse and abuse of drugs, especially suicide drug poisoning. Legislative actions may help in controlling drug use among adolescents and young adults and preventing them from getting access to specific drugs.

Mensah et al. [31] investigated factors leading to addiction and abuse of opioids. Prince et al. [8] found that there is a direct connection between drug use and mental illness. People who have a record of past hospitalization or using prescribed painkillers for mental illnesses (e.g. schizophrenia, bipolar disorder, major depressive disorder) are more likely to seek drugs afterwards. Suicide attempts among people with severe mental illness (SMI), who used prescribed painkillers, were 2.40 times higher than for people with other substance use disorders. People with SMI are more likely to have an opioid use disorder (OUD) and OUD was present in 5% of people with SMI. However, they were unable to identify the neurobiological risk factors in their Syndrome Model. A related study [9] presents that people, who have a record of OUD, are also likely to have mental disorders and to use other substances (nicotine, alcohol, tranquilizers, etc.). We may conclude that there is a connection between mental illness and drug use. However, there is a correlation, causality is less clear. Van et al. [10] identified the connection between mental disorders and opioid overdoses. In their review study, authors tried to explain

the association between mood/depression disorder and opioid overdose. There also is an association between PTSD (post-traumatic stress disorder) and opioid overdoses.

Bohnert et al. [11] found OUD has a strong connection with suicide attempts and overdose deaths. About 40% of suicide and overdose deaths were related to opioid use disorders in 2017 in America. Some factors cause people to use opioids in order to cope with society but this may increase depression, stress, anxiety, pain and eventually lead to suicide or overdose death. Easy availability of opioids and the use of other substances combined with opioids further increases the risk of unintentional overdose deaths.

Table 2.1: Comparative Analysis Table on Related Opioid Research

Category	Contribution	Limitation
Intentional and Unintentional Overdose	Comparison was made between intentional and unintentional cases [30]	The number of cases of exposure and deaths was not included in the analysis due to the lack of proper information.
Intentional and Unintentional Overdose	Identified pattern of intentional opioid drugs by sex, age, season and day of the week [7]	Only a certain age group is represented which may leads to biased results
Intentional and Unintentional Overdose	Describe the trends in misuse of benzodiazepines, and opioid analgesics [14]	The rising incidence of newborns experiencing withdrawal syndrome was not discussed
Intentional and Unintentional Overdose	Investigate the reason for unintentional poisoning [32]	Lack of causality and cross validation was not properly done in thier small cohort of data
Overdose linked to mental health	A brief systamatic review on mental linkage to opioid overdose [10]	The studies reviewed in this paper didn't have any experimental study analyses
Overdose linked to mental health	Finding association of overdose with Mental Health [Angoli]	The analysis was done from the predetermined data and there was a lack of determination of specificity and sensitivity of this measurement as well as the reasons why dose tapering associates mental health crisis and suicidal attempt were not discussed here.
Overdose linked to mental health	The research concentrated on the linkage between drug use and mental disease [8]	There was a scarcity of reliable evidence on why attempted suicide is increased in those who use prescription painkillers.
Overdose linked to mental health	According to the author, either mentally ill individuals use drugs or drugs make them mentally ill. [9]	Lack of causality or prevention policy
Overdose linked to mental health	Understanding Links among Opioid Use, Overdose, and Suicide[Novak]	The survey presented in this paper was done on self-reported opinions and the opinions were cross-sectional
Overdose linked to mental health	Discussed on factors associated with opioid overdose and prevention starategy [[11]	This paper has no statistical analysis and only used predetermined data.

Several studies [14, 32] have attempted to predict opioid abuse using traditional statistical approaches. Alzeer et al. [33] wrote a review paper attempting to find crucial indicators from literature. Their study identified 75 factors that are connected with opioid abuse. They found age and gender to be the most important indicators. Vunikili [13] also presented a set of statistical models to classify patients at risk of opioid misuse, death, and drug-drug interactions. Machine learning is an emerging field for predictive analysis and is used in multiple

sectors. Han et al. [12] developed a prediction model to demonstrate the efficacy of machine learning in predicting opioid patients. The MIMIC-III dataset was used to classify opioid-dependent patients.

In this study, we consider the social determinants and the patient characteristics, particularly using the structured set of data from the MIMIC-III database. In addition, making use of the unstructured data could improve the decision-making process. To the best of our knowledge, no study has investigated the interaction of behavioral health and opioid drugs using domain-specific word embedding. The contribution of our research is to classify opioid patients from structured as well as unstructured data.

It is challenging to classify opioid patients with unstructured data using machine learning algorithms. However, it may be feasible with the aid of knowledge distillation. Ahn et al. [34] introduced the concept of the knowledge distillation which was formulated by Hinton et al. [35]. According to Ahn et al. [34] knowledge from a higher capacity model could be compressed to a lower capacity model by training the weaker model with the logits generated by the stronger model. However, it has been established in the research by Gao et al. [36] that it is feasible to distill dark knowledge from a totally different presentation of data from a strong network to a weaker network.

2.3 Gap Analysis

In this study, we considered the social determinants to identify the characteristics of a patient, particularly from the structured set of data from the MIMIC-III database. Additionally, the unstructured portion would bolster the decision-making process. To the best of our knowledge, no study has investigated behavioral health with opioid drugs and used domain-specific word embedding. so, the vital step in our research is to classify opiate patients both from structured and unstructured data.

Chapter 3

Methodology

We have divided this section into four subsections. Begin with the dataset description; we discussed the dataset. Then, we conducted a search strategy with the opioid-related keyword and from various websites and papers. Afterward, case identification was carried out for the respective patients using the keywords we collected earlier. In this problem, we have input patients' characteristics, i.e., ethnicity, demography, and lab value. Other inputs are the doctor's prescription and ICD-9 codes. The output is whether or not a patient is opioid-addicted; if addicted, is it an intentional or unintentional case.

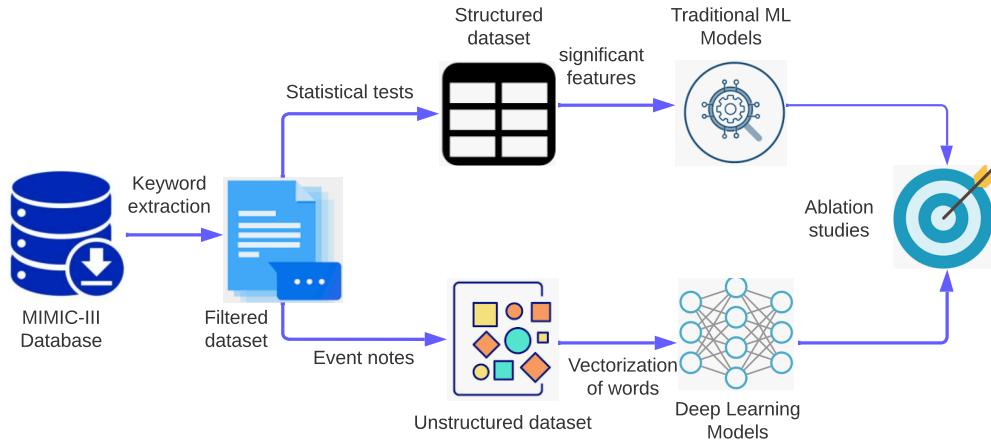


Figure 3.1: Methodology of opioid misuse prediction from

3.1 Data Description and Statistics

The MIMIC III database [37] has 26 schemas but only ten schemas are relevant to our study. To create our dataset, 41 relevant features were taken. Within the unstructured, data, we identified a total of 37,127 distinct cases which we filtered systematically. We extracted our cohorts from the MIMIC-III database. MIMIC-III is a massive, publicly available database which contains health-related information of over 40,000 patients from the Beth Israel Deaconess Medical Centre's intensive care units during the period from 2001 to 2012. Demographics, vital sign assessments taken at the bedside, diagnostic test findings, treatments, prescriptions, caregiver observations, imaging documents, and death details are also stored (both in and out of the infirmary). The database has information on 53,423 different admitted patients (aged 16 years and older) in critical care units. Furthermore, it encompasses information of 7,870 neonates who were admitted between 2001 and 2008. The information comprises 49,785 children and 38,597 adult patients.

The tables in MIMIC-III [18] are connected by identifiers that ordinarily end in “ID”. For instance, the SUBJECT ID is a single sufferer, while the HADM ID signifies a hospital admittance, and the ICUSTAY ID signifies an admittance to an ICU. A descriptions of MIMIC-III is summarized in Table 3.1.

Table 3.1: Statistics of the MIMIC-III database.

Dataset features	Total
Total Patient	46,520
Distinct Hospital Admission	49,785
Number of schemas	26
Number of distinct ethnicity	42
ICD-9 based Opioid cases	212
Median Age group	65.9
Gender (Male)	55.9%
Gender (Female)	44.1%
ICU mortality	8.5%

3.2 Dataset Preparation

In this section, we prepare our training dataset through different steps. First we preprocess our dataset, then we annotate our dataset by domain experts by ICD-9 codes. Finally, we use the dataset to train our model.

3.2.1 Pre-processing

For the two types of datasets, several data preprocessing techniques have been used. For the structured (i.e., tabular) dataset, the label encoding technique has been used initially since many of the attributes had qualitative data. The tabular dataset had a few missing values as well. There are several missing value handling techniques which are available to deal with this, for example, discarding the instance, replacing with mean/mode/median, replacing with next/previous value, imputing the most frequent value, KNNImputer, etc. Among those techniques, KNNImputer [38] was selected for handling the missing values. KNNImputer is widely used for missing value handling [39, 40] and provides the best outcomes for the tabular dataset that we are working with. Figure 3.2 shows different techniques of missing values imputation.

3.2.2 ICD-9 codes

The World Health Organization (WHO) developed the International Classification of Diseases (ICD) as a standard diagnostic tool for evaluating the incidence and prevalence of illnesses and other ailments. In this section, we referred to ICD-9 codes from the MIMIC-III and only included samples with distinct ICD-9 codes indicating opioid cases.

With ICD-9 based Opioid code, We have identified a total 212 patients who possessed mental and social conditions directly linked to Behavioral health. Also, the following tables show the ICD-9 code that we have

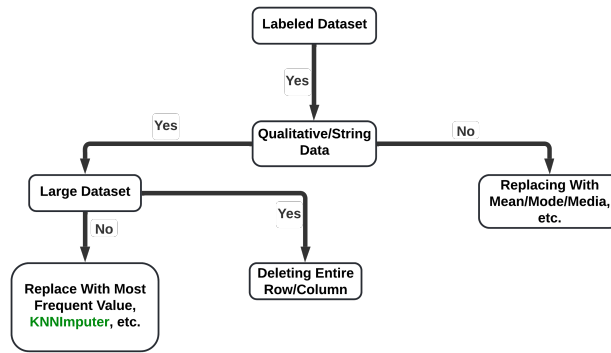


Figure 3.2: Missing Value handling technique selection

used to find the correlation between drug dependency and mental illness. Table 3.2 shows ICD-9 codes and its descriptions.

Table 3.2: Association with Behavioral Health and Opioid Intake

Subject_id	ICD-9 code and it's meaning	Types of Opioid intake	Behavioral Health
109	304.00,Opioid type dependence	Morphine, Vicodin	hypertensive crisis
636	305.52, Opioid abuse	Morphine	Systolic hypertension, bipolar disorder
1860	965.02, Poisoning by methadone	Methadone	Chronic Pain, Divorced
6287	E850.0, Accidental poisoning by heroin	Heroin	Depression, Lives alone
7450	304.02, Opioid type dependence	Cocaine, Morphine	Alcohol abuse, Rehabilitation
21156	305.51, Opioid abuse	Fentanyl	Unemployed, Depression

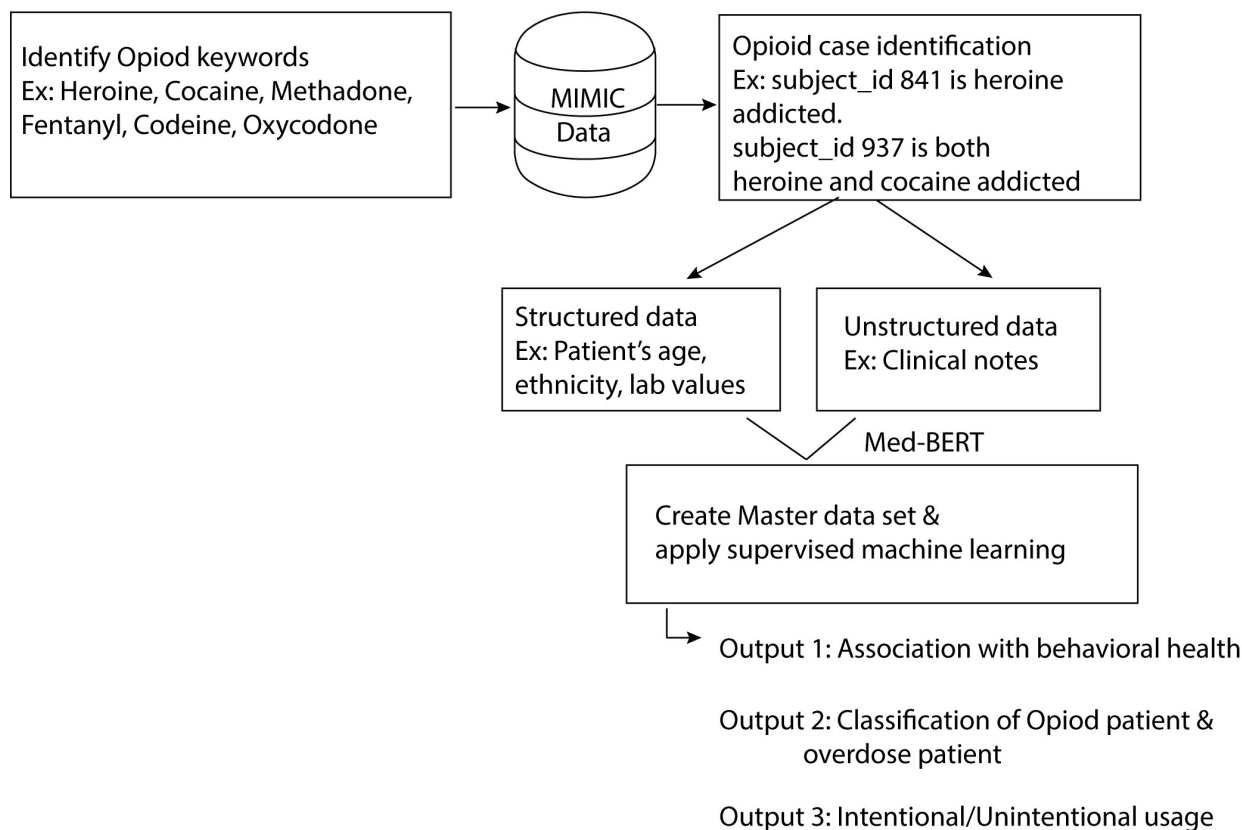


Figure 3.3: Keyword base steps of Predicting Intentional/Unintentional Opioid Cases

3.2.3 Ground truth

We manually selected data from patient's prescriptions for the ground truth collection. We asked the following questions: *is the patient taken any narcotics, did the patient take any opioids, did the patient use any controlled substances?* After a careful review, we selected 500 samples of opioid overdose and mental health conditions like *depression, hypertension, bipolar disorder*. It is important to mention that MIMIC-III is a publicly available data source for research purposes, and we only rely on this source. We have not used or amalgamated any other external data. Furthermore, no personally identifiable information (PII) was disclosed, and HIPAA policy was maintained during ground truth collection.

We got the text column from the 'Noteevents' table, where the prescription of the patient is stored. The *Noteevents* is the only table in the MIMIC-III dataset which contains all the notes of the patients, and is a comprehensive source of the unstructured data. Each note is linked to the specific subject id of the patient containing information about admission type, past medical history, socio-economic status, and a detailed description of the patient. For each patient, we searched for any opioid related information in the prescription. As mentioned earlier, there is specific information for each patient in all the prescriptions such as *past medical history, social history, and family history* along with the medication which is provided to the patients. We searched manually in

the prescriptions to find any evidence of opioid misuse from the *past medical history* and we also search for information on depression, anxiety, living alone, or broken family that can potentially cause someone to feel unhappy, in the ‘Social History’ and ‘Family History’ categories. For example: we identified a patient (patient id: 7445) as opioid dependent or potentially a future abuser.

According to the objectives of this study, we develop datasets containing the socio-economic characteristic of the patients, in addition to the related factors like lab events and vital signs, to identify opioid patients. We performed a qualitative study by doing a brief literature search select only schemas and their relevant columns which may be helpful for decision making. After several data preprocessing steps, high level categorization, and feature engineering, we found 41 factors useful for this study. Table 3.3 shows the selected tables and their columns to form a structured dataset for our study.

Table 3.3: Statistics of our filtered dataset.

Dataset features	Total
Number of keywords (structured)	121
Number of keywords (unstructured)	128
Number of features	41
Number of patients	454
Intentional cases	182
Unintentional cases	272
Age (median)	38 years
Gender	M=207 and F=247
Length of stay (median)	48 days

The dataset that we created for the analysis has a shape of (454*41) which we feed into the model to classify intentional and unintentional opioid patients. As mentioned in our contribution section, this dataset is made publicly available for further research work.

3.2.4 Agreement Analysis among Annotators

For our data analysis purposes, we convert the ICD9 codes into higher level disease categories. We perform this task to produce a lower number of class labels. According to the original dataset, if we use each ICD9 code as a single class label there is a high chance of overlap in the class prediction for the feature. For example, respiratory system disorder and shortness of breathing. We did not consider these diseases in two different classes, rather we lower down the class label as respiratory issues. In this way, we convert the 51 independent class labels to only eight class labels.

To this end, we recruited three physicians from Directorate General of Health Services (DGHS), Bangladesh (01), and Dhaka Medical College Hospital, Bangladesh (02). We hired three physicians as annotators of the ICD9 codes, because converting the result according to a majority voting scheme. First, we produce the

same spreadsheet in three independent copies. We confirmed that these annotators performed the task in an independent fashion, so that one annotator does not get influenced by the other annotators.

Since we have three annotators, we use Fleiss' kappa [41] to compute the final class label. To what extent raters assign the similar score is called agreement in their observation. To obtain a fair percentage of agreement, statisticians create a matrix where column and row represent raters and objects that are going to be rated, respectively. If two raters have complete agreement between their ratings, then we consider a zero (i.e., no disagreement), else we assign a one which represents a disagreement. Then, we find the percentage of zeros which represents the agreement score, i.e., Kappa [41]. Kappa can range from -1 to +1. After calculating the agreement score by using Fleiss' kappa, we obtain kappa scores of 0.9 among the physicians which indicates strong agreement.

3.2.5 Case Identification Based on Keywords

In this step, we selected 121 opioid-related keywords (i.e., Hydrocodone, Methadone, Fentanyl, etc.) from the literature [42–44]. We verified whether the opioid-related keywords referred to prescription opioids or to illegal addictions in this identification process. This qualitative search helps us to remove irrelevant keywords. As a result, we identified 32,152 opioid-dependent patients from the prescription tables of MIMIC-III database with selected keywords. The prescriptions schema is about the medications ordered for a given patient. The opioid keywords we chose performed detailed queries by finding one or more opioid drugs served to the patients and returned distinct identifiers as subject id. The purpose of our initial data preprocessing was essentially to find the opioid related patients from the MIMIC-III database, and the group of patients who were given treatment for an overdose. As MIMIC-III is a huge dataset of patients with various diseases, and treatments, so we first segregate the opioid patient cohort, and to include the maximum numbers of patients. We applied every possible opioid related keyword where we could find out of the 26 tables and 202 features, we found 41 features that are relevant to our research. Therefore, the shape of our initial dataset was (32,152*41).

we have identified opioid-dependent patients based on the Opioid Drug List Table. So, we have identified cases based on the keyword generated by us. In that regard, we have created a master file named "MIMIC" by gathering the relevant data from a total of twenty-six schemas housed in the MIMIC database. As a result, we can identify more opioid-dependent patients; and, any unidentified patients can now be recognized with the analysis. Also, to the best of our knowledge, no one in this area has explored the potential of binding the unstructured data and ICD-9 codes with patients' biological measures and associated social determinants.

3.3 Model Building

We use a range of approaches, such as tabular data from correlating variables and unstructured data from patients' history notes, to create models to predict opioid patients. The classification of opioid patients from both structured and unstructured datasets has been done using a variety of classification algorithms and methods. We describe the detailed steps in the following chapter regarding model building.

Chapter 4

Model building with Ablation Studies

4.0.1 Model Building

We build models to predict opioid patients using two different approaches: i) tabular data from correlated attributes and ii) unstructured data from patients' history notes, i.e., eventnotes. In this subsection, we first explain the method of building opioid intake behavior prediction by using the tabular data. Later, we also describe the process of model building by using our unstructured data.

Table 4.1: Performance of different classifiers to predict opioid behavior by using structured dataset.

Model (Classifier)	Training Accuracy	Testing Accuracy	Training MCC	Testing MCC	Training Sensitivity	Testing Sensitivity	Training Specificity	Testing Specificity
AdaBoost	0.953	0.923	0.906	0.847	0.953	0.950	0.950	0.900
Logistic Regression	0.956	0.934	0.912	0.868	0.958	0.950	0.953	0.922
Support Vector	0.956	0.934	0.912	0.868	0.958	0.950	0.953	0.922
XGB	0.956	0.934	0.917	0.868	0.958	0.950	0.959	0.922
Random Forest	0.967	0.934	0.934	0.868	0.964	0.950	0.971	0.922
ANN	0.959	0.959	0.916	0.956	0.975	1.0	0.938	0.960

Structured data classification

We build our structured dataset, D_t , with the correlated attributes: mental status, ethnicity and ICD high level category (diagnosis). We then split our dataset (instances 454) into training and testing datasets of 80% and 20%, respectively. We train our dataset with a cross validation with 10-iterations by using the following classifiers [45]: AdaBoost Classifier, Logistic Regression, Support Vector Classifier, XGB Classifier, and Random Forest Classifier. Table 4.1 shows the performance of our models which are developed from the tabular dataset.

Unstructured data classification

For unstructured data classification, where the patients' histories are the input of the models, several classification approaches can be used. Since the unstructured data do not have any particular format, our study performs the classification by NLP techniques. The unstructured data consists of two attributes where the first attribute refers to the patients' history and the second attribute is our target attribute opioid intake intentional YES/NO.

We applied several preprocessing techniques to our unstructured dataset. The data cleaning technique is used to remove rows of incomplete data from the dataset. We discard a few irrelevant syntaxes such as brackets and punctuations from patient's history data field. We also replace some abbreviations (e.g., dr./Dr./md./m.d.) with their full form so that the words do not appear as individual sentences at the time of vectorization. Word vectorization [46] is an NLP method of mapping a word or phrase from a vocabulary to a corresponding vector. The unstructured data, D_{us} , has been classified using three different well known methods: 1D CNN-based model, a basic LSTM-based model, and an LSTM and attention-based model.

LSTM and Attention techniques: Long Short Term Memory Networks, most commonly referred to as "LSTM" are a unique class of RNN that can recognize long-term dependencies. When there is a lot of information to summarize, the model performs poorly and produces inaccurate results. It is known as the RNN or LSTM long-range dependency problem. The attention mechanism with the LSTM attempts to address the problem. The cell state is the foundation to LSTMs. The LSTM can modify the cell state by removing or adding information, which is carefully controlled via gates. Information can pass through gates voluntarily. These three gates serve to preserve and regulate the cell state in an LSTM architecture. Choosing whatever information from the cell state to discard is the first stage in our LSTM model. The forget gate layer, which is on sigmoid function, decides what to discard. Equation 4.1 shows how the forget gate layer scans x_t and h_{t-1} :

$$f_t = \sigma(W_f[h_{t-1}; x_t] + b_f) \quad (4.1)$$

The next step is to choose the new information that will be kept in the cell state. Two different components perform this task. The input gate layer i_t , a sigmoid layer, first determines which values will be updated. The state is then updated with a vector of potential new values, $\tilde{c}$, created by a tanh layer. These two will be combined in the subsequent phase to produce an update to the state as follows according to

$$i_t = \sigma(W_i[h_{t-1}; x_t] + b_i) \quad (4.2)$$

$$\tilde{c}_t = \tanh(W_c[h_{t-1}; x_t] + b_c) \quad (4.3)$$

In the very next stage, we update the old state. We multiply the previous condition by f_t while omitting the earlier items on our list of information to forget. Then $i_t \tilde{c}_t$ is added and it makes a new candidate value which is presented in Equation 4.4.

$$c_t = i_t \otimes \tilde{c}_t + f_t \otimes c_{t-1} \quad (4.4)$$

Based on the output of candidate state, we finalized the output of LSTM cell in the final output gate as follows (according to Equation 4.5):

$$o_t = \sigma(W_o[h_{t-1}; x_t] + b_o) \quad (4.5)$$

$$h_t = o_t \otimes \tanh(c_t) \quad (4.6)$$

In a conventional LSTM unit, the sequences are only encoded in one direction (past information). In order to preserve both past and future information, we employ a BiLSTM model [47] to encode the sequences in both directions.

Natural language processing, machine translation, and image processing tasks are successfully using the attention mechanism. In the supplied input sentence, it extracts pertinent context information about a word. The bidirectional LSTMs forward and backward output features are concatenated into the vectors h_t it before applying attention. The formulation of the attention mechanism is described in Equations 4.7, 4.8 and 4.9 :

$$\alpha_{ts} = \frac{\exp(\text{score}(h_t, \bar{h}_s))}{\sum_{s'=1}^S \exp(\text{score}(h_t, \bar{h}_{s'}))} \quad (4.7)$$

$$c_t = \sum_s \alpha_{ts} \bar{h}_s \quad (4.8)$$

$$a_t = f(c_t, h_t) = \tanh(W_c[c_t; h_t]) \quad (4.9)$$

$$\text{score}(h_t, \bar{h}_s) = v_\alpha^\top \tanh(W_1 h_t + W_2 \bar{h}_s) \quad (4.10)$$

In a global attention model, all of the encoder's hidden states are taken into account when determining the context vector c_t . In this model type, the current target hidden state h_t and all source hidden states $\bar{h}_s$ are compared to derive a variable-length alignment vector a_t , whose size is equal to the number of steps (according to Equation 4.10). The weighted average of all the source states, according to a_t s, are then calculated to create a global context vector, or c_t . Therefore, the computation path of this attention mechanism is, calculation h_t ,

then the attention weights a_t s, then the context vector c_t and finally the attention vector a_t .

To learn the text's sequence, the LSTM-based model contains one LSTM layer with 400 units. This model has 3.1 million parameters, making it more complicated than the other models. Additionally, a hybrid (LSTM + Attention) model was employed to classify the sequence. The attention layer, which outputs a 128-dimensional vector, comes after a bi-directional LSTM layer with 128 units for each layer. For each of the models, there is a dense layer that is fully connected and has two units. The output layer includes two units since the model requires a logit to pass through the softmax function. Adam serves as the optimizer for all models, with the same learning rate (0.001), batch size (32) and Data split (80/20).

1D CNN: 1D CNN is the modified version of the 2D CNN. It is widely used in sequence learning which has a shallow layer. Therefore, it uses minimum resources to train and test. Compact 1D CNNs have shown improved performance in recent research for applications with low labeled data and high signal variations obtained from various sources [48].

We use a non causal system of CNN since the output y is dependent on a future sequence of inputs x . Let the input to convolution layer of length n be represented by x , and let the kernel of length k be represented by h . Let the kernel window be shifted s positions (number of strides) after each convolution operation. Then a non-causal convolution between x and h for stride s can be defined as follows:

$$y(n) = \sum_{i=0}^{LK} x(n + i + (s1))h(i) \quad (2)$$

Table 4.2: Summary of the models that have been used to classify unstructured dataset.

Network	1d CNN	LSTM	LSTM+Attention
Architecture	Embedding, 128 1d CNN, 32 Dense, 2	Embedding, 128 LSTM , 400 Dense 2	Embedding, 128 BiLSTM, 128 Attention , 128 Dense, 2
Total Params	2.3M	3.1M	2.4M
Iteration / Batch	11/32	11/32	11/32
Optimizer	Adam 1e-3	Adam 1e-3	Adam 1e-3

In our 1-D CNN model. we use a 128-dimensional embedding layer and makes use of 32 kernels in total. In terms of parameters, the model is less complex than the others. Starting with a layer of embedding of 128 dimensions, all models are constructed.

4.0.2 Combining models by Knowledge Distillation

Given an eventnote x from a dataset D_{us} , consisting of patients history in texts with corresponding labels where D_{us} is an unstructured dataset. The note consists of $t_1, t_2, \dots, t_n$ tokens which are extracted from the texts. Our

aim is to train a model M_S on the dataset D_{us} such that it achieves an accuracy a by using distill knowledge from a model M_T . The model M_T is trained on structured dataset D_t with limited features $t_1, t_2, \dots, t_k$, where $n \gg k$.

We propose a novel method to classify misused opioid patient's from unstructured data, D_{us} by using the knowledge of a model that is trained on structured data. Figure 4.1 presents the details pipeline how we combine both structured and unstructured datasets by using knowledge distillation from the structured dataset. In our knowledge distillation approach, a model is trained on unstructured data, D_{us} , learns from the insights extracted from a structured data, through a shallow ANN (Artificial Neural Network) model. In this case, the teacher model, M_T , is trained on dataset, D_t , which is a tabular structured dataset with limited features. The dataset D_t is constructed by the domain experts (see Section ??). Therefore, the dataset D_t is trustworthy and the model is learning the known limited features while training. The main purpose of the teacher model (M_T) is to provide insights on the set of unstructured training set (D_{us}) of the student model (M_S).

Initially, We look at the label of dataset D_{us} , and we extract features from dataset D_t for the corresponding label, and created a new dataset. Therefore, we get a new dataset D_{sn} whose label's is identical to the label of dataset D_{us} . The student model outputs two-dimensional score vectors for each input. Additionally, the teacher model gives us scores that resemble the output of the student model. Now, these can be utilized to calculate soft probabilities. To soften probabilities, We make use of the hyper-parameter temperature τ . When $\tau = 1$, softmax produces its typical output. However, when we raise, the softmax output softens and reveals which classes our teacher model discovered to be more similar to the predicted class. Hinton et al. [35] called it dark knowledge. The teacher model itself implant the dark knowledge during training. However, during the distillation process, this dark knowledge is transmitted to the student model which is built from unstructured dataset, D_{us} . According to the experiments of the authors [35], the value of τ could be from 1 to 20. Authors find that the same value of τ to the student and teacher models likely return the maximum results.

Formally, Let (x, Y) in D_{us} where x is an eventnote and Y is the corresponding label. Now our student model M_S , given an input x will output logits L_S which can be shown as $L_S = M_S(x)$. These logit values are softened by using the temperature τ and used in the softmax function σ to get the soft probabilities denoted by $\hat{Y}_{S\tau} = \sigma(L_S/\tau)$.

On the other hand, Y_S denotes the hard probabilities in $Y_S = \sigma(L_S)$ to be used by the CE (cross entropy) loss.

The teacher model M_T , outputs the score for each inputs from dataset D_{sn} . Assuming, (I, Y) in D_{us} where i is the set of structured features, gathered from the dataset D_s and Y is the corresponding label that is identical to the label of dataset D_{us} . Therefore, The score provided by the model, can be denoted as $L_{T_i} = M_T(I_i)$ and the hard probability distribution for each input can be shown as $Y_{T_i} = \sigma(L_{T_i}/\tau)$, the soften probability would be then, in $\hat{Y}_{T\tau} = \sigma(L_T/\tau)$. The final loss function now can be derived as Equation 2:

$$Loss_{kd} = (1 - \alpha)CE(Y_S, y) + (\tau^2 + \alpha)KL(\hat{Y}_{S\tau}, Y_{S\tau}) \quad (2)$$

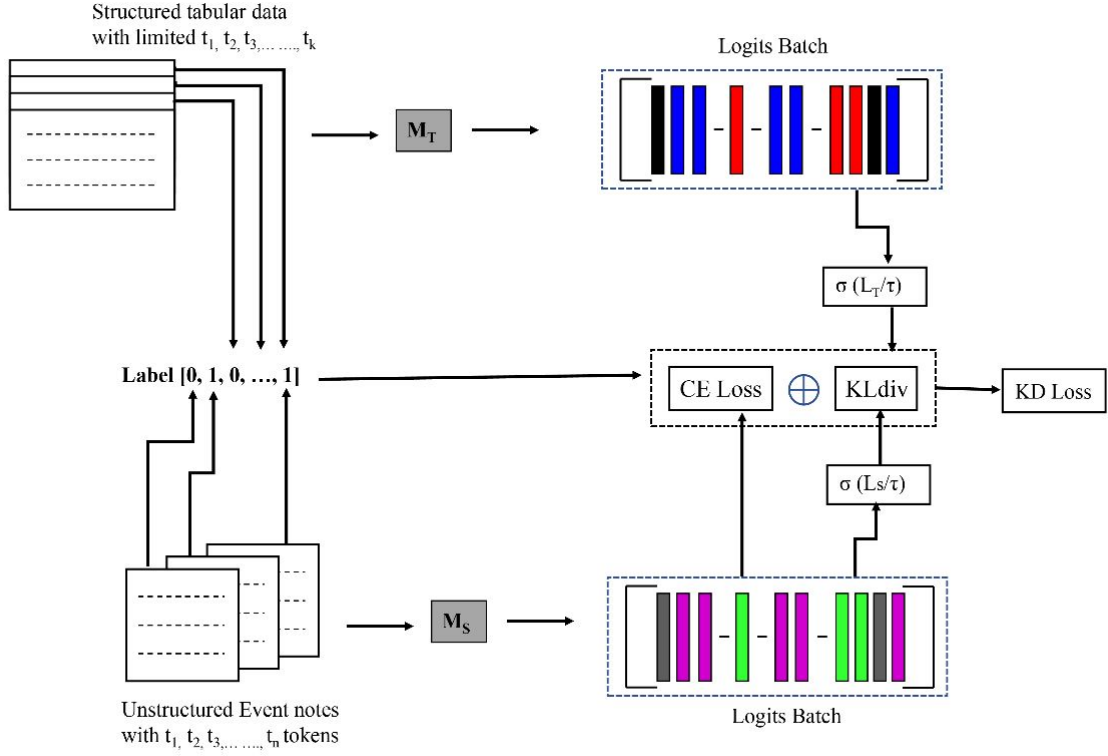


Figure 4.1: Combining models (structured+unstructured) by knowledge distillation.

4.1 Results

Several classification algorithms and techniques have been used to classify the opioid patients from both structured and unstructured datasets.

4.1.1 From structured dataset

In the comparison of application of ANN (Artificial Neural Network) and traditional machine learning algorithms for classification of tabular dataset different values of accuracy have been achieved. Table 4.1 presents the results of the models that have been used to classify the tabular dataset. The traditional machine learning classification algorithms such as AdaBoost, Logistic regression, Support vector, XGB and Random Forest provide a training accuracy of 95.3% - 96.7% and testing accuracy of 92.3% - 93.4%. Random Forest classification algorithm gives the best outcome in-terms of both training and testing accuracy among the traditional machine learning algorithms. On the other hand, the ANN algorithm provides a training accuracy of 95.9% and testing accuracy of 95.9%. Comparing ANN with traditional classification algorithm's we found that ANN provides slightly better testing accuracy in terms of classifying the tabular dataset.

4.1.2 From unstructured data

Several machine learning algorithms and techniques have been also used to classify unstructured dataset and the several values of accuracy have been achieved. Table 4.3 represents the results of different classification algorithms by using different approaches to classify the unstructured opioid patients. Here the results can be divided into two approaches:

1. Classification without using KD.
2. Classification using KD.

In terms of classification of unstructured data without using KD (knowledge distillation), 1D-CNN, LSTM and Hybrid (LSTM + Attention) models have been achieved testing accuracy of 64.8%, 54.9% and 61.0%, respectively.

By using the KD technique the outcomes have been improved for the same classification algorithms. The testing accuracy for the unstructured data classification using KD technique are 65.9%, 57.2% and 76.44% for 1D-CNN, LSTM and Hybrid (LSTM + Attention) respectively. 1D-CNN has been performed better than other algorithms in both approaches.

Table 4.3: Performance of different classifiers to predict opioid behavior from unstructured eventnotes.

	1d CNN	LSTM	LSTM+Attention
No KD	64.835	54.94	61.044
KD	65.93	57.23	76.44

4.2 Ablation study

A component of a machine learning architecture may typically be deleted or replaced as part of an experiment called an ablation study [49] to determine how these changes affect the overall performance of the system. The performance of a model may remain stable, improve or get worse when these components are changed. The accuracy can be improved by experimenting with various hyper-parameters like optimizers, learning rates, loss functions and batch sizes. Altering the model's architecture has an effect on overall performance as well. In this study, our suggested model is examined by arbitrarily removing or changing various components and parameters.

4.2.1 Ablation study 1: changing hidden layers

Between the input and output layers is a layer known as the hidden layer, where artificial neurons receive a series of weighted inputs and generate an output using an activation function. The performance of the model

is influenced by the hidden layers. Arbitrarily, we chose a single dense layer, which is a dense output layer. We observed a considerable change in the CNN and attention-based model results if we increased the number of hidden layers. The accuracy of the identical model with three hidden layers LSTM based mode, however, remains the same, although it significantly alters the training accuracy. Table 4.4 presents the performance of different models for different numbers of hidden layers. The table presents that our models become overfitted. The main reason for the weak performance due to small training dataset. The size of our selected patients' data is only 212, because these patients have association with mental health issues. We plan to find more patients information who are linked with mental health issues.

Table 4.4: Performance measured by changing hidden layers.

Hidden Layers	Model (Classifier)	Training Accuracy	Testing Accuracy	Findings
1	LSTM	89.47%	56.04%	Identical accuracy
2		47.64%	53.84%	Identical dropped
3		59.11%	56.04%	Identical accuracy
1	1D-CNN	100%	62.63%	Identical accuracy
2		100%	48.35%	Accuracy dropped
3		98.33%	59.34%	Accuracy dropped
1	Attention	99.1%	58.2%	Identical accuracy
2		97.78%	57.14%	Accuracy dropped
3		93.91%	54.94%	Accuracy dropped

4.2.2 Ablation study 2: changing batch size

The number of training samples used in one iteration is referred to as the batch size. If fewer samples are used to train the network, it uses less memory in the process overall. Mini-batches typically help networks train more quickly. We do so because the weights are updated following each propagation.

Experimenting with fewer samples shows that a batch size 32 appears to be optimal for all three models. We observe that changing the batch size can reduce test accuracy. Table 4.4 presents the performance of different models with different batch sizes.

4.2.3 Ablation study 3: changing optimizer

We used different optimizers to investigate the performance of our models. We found that the Adam optimizer performed the best among all the optimizers. For all three optimizers, we employ the same learning rate and loss function. For this dataset, SGD [50] did not perform well and RMSprop [51] did not outperform the Adam optimizer. Table 4.4 represents the performance of different optimizers for the models.

Table 4.5: Performance Measured by Changing Batch Size

Batch Size	Model (Classifier)	Training Accuracy	Testing Accuracy	Findings
16	LSTM	89.47%	53.84%	Accuracy dropped
32		89.47%	56.04%	Identical accuracy
64		60.11%	49.45%	Accuracy dropped
16	1D-CNN	91.96%	56.04%	Accuracy dropped
32		100%	62.63%	Identical accuracy
64		100.0%	57.14%	Accuracy dropped
16	Attention	99.7%	53.8%	Accuracy dropped
32		99.1%	58.2%	Identical accuracy
64		100%	51.6%	Accuracy dropped

Table 4.6: Performance Measured by Changing Optimizer

Optimizer	Model (Classifier)	Training Accuracy	Testing Accuracy	Findings
SGD	LSTM	51.2%	56%	Accuracy dropped
Adam		89.47%	56.04%	Identical accuracy
RMSProp		52.35%	46.15%	Accuracy dropped
SGD	1D-CNN	47.36%	54.94%	Accuracy Dropped
Adam		100%	62.63%	Identical accuracy
RMSProp		100.0%	62.63%	No Change
SGD	Attention	52.6%	47.2%	Accuracy dropped
Adam		100%	58.2%	Identical accuracy
RMSProp		100%	56%	Accuracy dropped

4.2.4 Ablation study 4: changing learning rate

The learning rate [52] indicates how frequently the weights are updated during training. The learning rate is a hyperparameter that can be customized and is used to train neural networks. Its value is typically small and positive in the range of 0.0 and 1.0. The learning rate significantly impacts our models performance. For majority of the models, 0.01 is the best learning rate. However, ‘the accuracy increased for an attention-based model when the learning rate was 0.1 or 0.001. With such modification in learning rate, the performance is improved. Table 4.7 shows the performance of the models by using different learning rates.

4.2.5 Ablation study 5: changing dropouts

A method to prevent neural networks from overfitting is dropout regularization [53]. Dropout disables neurons and their associated connections at random. This step may change all the neurons to develop their generalization skills and keep the network from relying too much on individual neurons.

Table 4.7: A method to prevent neural networks from overfitting is dropout regularization on Performance measured by changing learning rates.

Learning Rate	Model (Classifier)	Training Accuracy	Testing Accuracy	Findings
0.1	LSTM	52.35%	46.15%	Accuracy dropped
0.01		89.47%	56.04%	Identical accuracy
0.001		56.23%	49.45%	Accuracy dropped
0.1	1D-CNN	98.33%	59.34%	Accuracy dropped
0.01		100%	62.63%	Identical accuracy
0.001		100%	53.84%	Accuracy dropped
0.1	Attention	100.0%	60.4%	Accuracy improved
0.01		100.0%	58.2%	Identical accuracy
0.001		100.0%	59.34%	Accuracy improved

We employ dropout in primary layers like LSTM and CNN because our identical models have only one dense layer. When we do not use dropouts in an LSTM-based model, the accuracy improves, but the accuracy declines to the same level when we apply more dropouts. The attention-based model is comparable in terms of test accuracy for close dropouts. For dropout 0.3, it functions a little better. However, if dropout is increased to 0.50 or more, it performs weaker. Table 4.8 shows the performance of our models using different drop out levels.

Table 4.8: Performance Measured by Changing dropout percentage

Dropout	Model (Classifier)	Training Accuracy	Testing Accuracy	Findings
0.0	LSTM	61.21%	57.14%	Accuracy increased
0.3		89.47%	56.04%	Identical accuracy
0.5		78.11%	56.04%	Identical accuracy
0.0	1D-CNN	98.33%	59.34%	Accuracy dropped
0.3		100%	62.63%	Identical accuracy
0.5		100%	53.84%	Accuracy dropped
0.0	Attention	100.0%	58.2%	Identical accuracy
0.3		98.34%	59.34%	Accuracy dropped
0.5		99.45%	54.94%	Accuracy dropped

4.2.6 Summary of the ablation study

Identical accuracy is the term that defines a result of a model with default hyperparameters. The accuracy only drops or increases when we change the hyper-parameters from the default. Thus, we labeled Accuracy dropped, Accuracy improved in the table.

We get the best accuracy for the 1d CNN based model when the optimizer is Adam, the batch size is 32, 1 hidden layer, the learning rate is 0.01 and the dropout rate is 0.3.

The LSTM-Based model performs well when we don't use any dropout, the learning rate is 0.01, the optimizer is Adam, the batch size is 32 and there is 1 hidden layer .

The Attention based model obtained 60.4% accuracy which is the maximum for this model. when the learning rate is 0.1 with no dropout layers, the optimizer is Adam, the batch size is 32, and there is 1 hidden layer.

4.2.7 Selecting the best model

In our study, we apply two approaches in our dataset as follows:

1. Classification without using KD.
2. Classification using KD.

The best model for approach 1 is the regular 1D CNN-based model since it has obtained the maximum accuracy among all the models applied. On the other hand, the gradient boosting classifier is the best model for approach 2. Approach 1 is a direct classification approach from raw event note data without assistance of a pre-trained model. The performance of the LSTM and the attention based models is comparable and cannot surpass the accuracy of the 1D CNN-based model. we obtained 62.63% accuracy in this case with a 1D CNN-based model. Approach 2 is an approach where we rely on a pre-trained neural model named Stanza. Here, we extract the features (test, problems, treatments) using the stanza model, relying on Stanza to train the model on the same data. In this case, the gradient boosting algorithm obtained the maximum accuracy of 74%.

Chapter 5

Discussion

In this chapter, we mainly discuss the major insights of our work.

Opioids are a class of drugs used for the relief of pain and many studies show that the use of opioids in USA is increasing day by day [54]. Opioid actually stops the pain signals between the brain and the body which may have long term consequences like addiction and even death. In our study, we find that ethnicity has a strong correlation with users' opioid behavior (see Table 2.1). We also find in a few studies that different ethnic groups have different number of opioid related deaths. In 2018-2019, the distribution was 73% nonHispanic White, 15% non-Hispanic Black, 7% Hispanic, and 6% other ethnicity communities [55]. A significant increase in death rate of around 38% had been observed for non-Hispanic Black individuals from 2018 to 2019, but there was no change overall among the other ethnic groups [55]. Other studies [56] found that there are some relationship between ethnicity and intentional behavior in terms of using opioid intake. Patients who belong to hispanic/latino, white Russian, white Brazilian, native Hawaiian, and other Pacific Islander, are likely 'No' opioid intentional intake behavior. On the other hand, Black/African Americans and few other patients have a larger rate of 'YES' opioid intake behavior.

Several studies [8–10, 57] show that there is a strong connection between users' mental health status and their opioid intake pattern. Among the 239.4 million U.S. adults, 38.6 million had a mental health disorder. Among the adults with mental health disorders, 18.7% are opioid users compared with only 5.0% among those without mental health disorders [57]. The study also shows that approximately 115 million opioid prescriptions are distributed each year in the US, 51.4% (60 million prescriptions) of which are received by adults who have a mental health disorder [57]. Our study is compatible with the previous findings. We also found that patients' mental health and intentionality of opioid using has a strong relationship. Almost every one of the patients who is suffering from depression, hypertension, and bipolar disorder has a tendency to misuse opioids intentionally. On the other hand, patients without mental health issues are not likely to misuse with opioid overdose.

Another interesting aspect is that mental health and social determinants are a controlling factor of drug abuse. Studies [58, 59] have shown that individuals with low levels of education and those who fall into high unemployment and poverty categories are at a greater risk of opioid abuse. Additionally, it suggests that people with higher socioeconomic status are prone to opioid abuse disorder than those with lower socioeconomic status. We used several models to classify the unstructured data. These models are: LSTM, CNN, and hybrid model

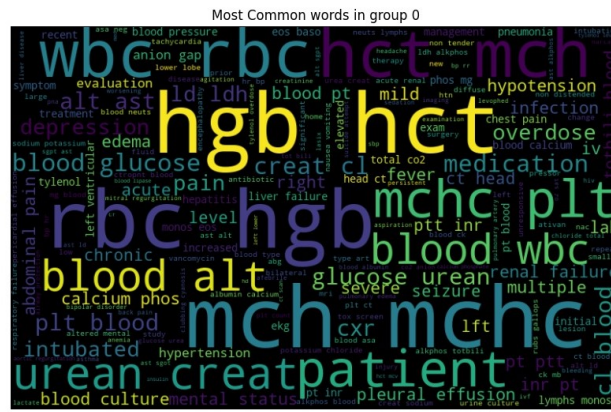


Figure 5.1: Frequently found problems, test and treatments name for unintentionally opioid users

to classify opioid intentional intake pattern. Among the three models, the CNN model is able to classify the unstructured data most accurately. The LSTM and Hybrid models need a larger dataset to train accurately and with only 453 patients' histories available, those two models obtained accuracies in the range of 52% - 54%. On the other hand, CNN is a comparatively simple model and it can understand the word sequence better in compare to other LSTM, therefore 1D CNN model was able to reach an accuracy of 64%.

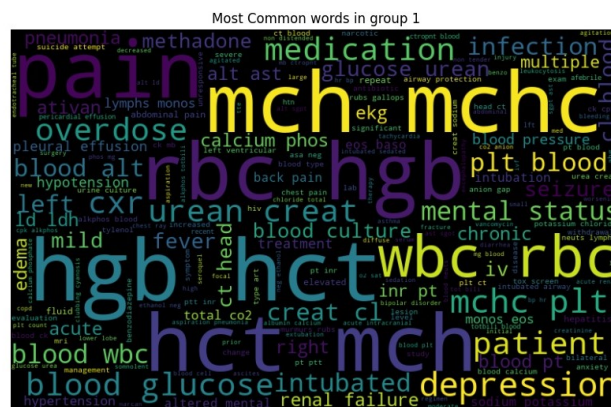


Figure 5.2: Frequently Found problems, test and treatments name for intentionally opioid users

Figures 5.1 and 5.2 show the patterns of word cloud in the history of patients who take opioid intentionally YES/NO. If we observe the problems, tests, and treatments for both classes, we see many familiar entities between the two categories. Therefore, we may find some overlap in the features of the two categories. However, we build a word cloud which is not an easy task to classify as they share almost the same features. Figure 5.1 shows that opioid intentional user No has an indication of mental health issues. On the other hand Figure 5.2 shows that opioid intentional YES users' have a strong tendency mental issues in their history such as depression, mental status, etc.

Our study has a number of shortcomings. We prepare independent models, but a combined model could improve the accuracy of the model. In our study, we identify that opioid usage has an association with users' mental

health issues. However, our models do not find which opioid has association with which mental health issues (i.e., depression, obsessive compulsive disorder, schizophrenia, etc.).

Chapter 6

Conclusion

Opioid use is a crisis globally among young and older people. In our study, we have built a dataset by using MIMIC-III dataset where we have narrowed down a total of 26 relational tables to only 41 features. We then identified correlated features in terms of opioid intentional YES/NO user. We identified three important features which have a strong correlation. In this way, we built a tabular dataset which demonstrated a good performance in predicting users' opioid intake behavior. We have also built a deep learning based model to predict users' opioid intake behavior from their historical information (event notes). By using our tabular model, we have obtained an accuracy of 93% by using random forest classifiers. Later, by using our deep learning (i.e., 1D CNN, LSTM+ attention), we have obtained an accuracy of 66% data from patients' unstructured historical data. After using the knowledge distillation mechanism of the tabular model over the deep learning based model, we have obtained an overall accuracy of 76.44%. We found some interesting correlations with users' mental health issues. There are a number of avenues to further improve our studies. We may increase the size of our dataset which might require more manual work to discover opioid intake YES users.

6.1 Limitation

One of the primary limitations of this project is the limited availability of structured data and the presence of unstructured data, which can be challenging to analyze. This may lead to an incomplete or inaccurate classification of opioid patients. Another limitation is the need for more robust modeling techniques that can handle complex and dynamic data sets, as well as more advanced feature selection methods. In addition, there is a need for more comprehensive evaluation metrics that can provide a more accurate assessment of the model's performance. Another limitation is the need for interpretability in the model, as it is essential to understand the features that contribute to the classification of opioid patients.

6.2 Future Work

One potential area of future work is exploring the use of knowledge distillation techniques to transfer knowledge from structured to unstructured data. This can leverage the vast corpus of unstructured data available to improve classification accuracy. Collecting more data, especially from diverse patient populations, can help improve the model's accuracy and robustness. There is also a significant opportunity to improve the interpretability of the model by using techniques such as feature importance analysis and visualization methods.

References

- [1] Aidan Maguire. Addressing america’s hidden killer: The opioid crisis.
- [2] National Institutes of Health et al. Opioid overdose crisis. national institute on drug abuse.
- [3] Christina Abdel Shaheed, Chris G Maher, Kylie A Williams, Richard Day, and Andrew J McLachlan. Efficacy, tolerability, and dose-dependent effects of opioid analgesics for low back pain: a systematic review and meta-analysis. *JAMA internal medicine*, 176(7):958–968, 2016.
- [4] Deborah Dowell, Tamara M Haegerich, and Roger Chou. Cdc guideline for prescribing opioids for chronic pain—united states, 2016. *Jama*, 315(15):1624–1645, 2016.
- [5] Jonathan Barkley and Jaejoon Shin. Analysis of intentional and unintentional drug overdose deaths occurring in rhode island, 2016–2019. *Rhode Island Medical Journal*, 105(1):46–48, 2022.
- [6] Beth Han, Wilson M Compton, Emily B Einstein, Jessica Cotto, Jennifer A Hobin, Jack B Stein, and Nora D Volkow. Intentional drug overdose deaths in the united states. *American Journal of Psychiatry*, 179(2):163–165, 2022.
- [7] Kelsey Pringle, Sarah Caupp, Junxin Shi, Krista K Wheeler, Henry A Spiller, Marcel J Casavant, and Henry Xiang. Analysis of intentional drug poisonings using ohio poison control center data, 2002–2014. *Clinical toxicology*, 55(7):652–658, 2017.
- [8] Jonathan D Prince. Correlates of opioid use disorders among people with severe mental illness in the united states. *Substance Use & Misuse*, 54(6):1024–1034, 2019.
- [9] Christopher M Jones and Elinore F McCance-Katz. Co-occurring substance use and mental disorders among adults with opioid use disorder. *Drug and Alcohol Dependence*, 197:78–82, 2019.
- [10] Jenna van Draanen, Christie Tsang, Sanjana Mitra, Vanessa Phuong, Arata Murakami, Mohammad Karamouzian, and Lindsey Richardson. Mental disorder and opioid overdose: a systematic review. *Social psychiatry and psychiatric epidemiology*, pages 1–25, 2021.
- [11] Amy SB Bohnert and Mark A Ilgen. Understanding links among opioid use, overdose, and suicide. *New England journal of medicine*, 380(1):71–79, 2019.
- [12] Dae-Hee Han, Shieun Lee, and Dong-Chul Seo. Using machine learning to predict opioid misuse among us adolescents. *Preventive medicine*, 130:105886, 2020.

- [13] MS Ramya Vunikili, Benjamin S Glicksberg, and Lakshminarayanan Subramanian. Predictive modeling of susceptibility to substance abuse, mortality and drug-drug interactions in opioid patients.
- [14] Robert G Carlson, Ramzi W Nahhas, Silvia S Martins, and Raminta Daniulaityte. Predictors of transition to heroin use among initially non-opioid dependent illicit pharmaceutical opioid users: A natural history study. *Drug and alcohol dependence*, 160:127–134, 2016.
- [15] Kun-Hsing Yu, Andrew L Beam, and Isaac S Kohane. Artificial intelligence in healthcare. *Nature biomedical engineering*, 2(10):719–731, 2018.
- [16] Fei Jiang, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, and Yongjun Wang. Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology*, 2(4), 2017.
- [17] Rohan Bhardwaj, Ankita R Nambiar, and Debojyoti Dutta. A study of machine learning in healthcare. In *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, volume 2, pages 236–241. IEEE, 2017.
- [18] Shirly Wang, Matthew BA McDermott, Geeticka Chauhan, Marzyeh Ghassemi, Michael C Hughes, and Tristan Naumann. Mimic-extract: A data extraction, preprocessing, and representation pipeline for mimic-iii. In *Proceedings of the ACM conference on health, inference, and learning*, pages 222–235, 2020.
- [19] Curtis Florence, Feijun Luo, and Ketra Rice. The economic burden of opioid use disorder and fatal opioid overdose in the united states, 2017. *Drug and alcohol dependence*, 218:108350, 2021.
- [20] Massimo Lusetti, Tatyana Ruzsics, Anne Göhring, Tanja Samardžić, and Elisabeth Stark. Encoder-decoder methods for text normalization. Association for Computational Linguistics, 2018.
- [21] Yang Liu, Yao Zhang, Yixin Wang, Feng Hou, Jin Yuan, Jiang Tian, Yang Zhang, Zhongchao Shi, Jianping Fan, and Zhiqiang He. A survey of visual transformers. *arXiv preprint arXiv:2111.06091*, 2021.
- [22] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [23] Kai Han, An Xiao, Enhua Wu, Jianyuan Guo, Chunjing Xu, and Yunhe Wang. Transformer in transformer. *Advances in Neural Information Processing Systems*, 34:15908–15919, 2021.

- [24] Alex Sherstinsky. Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D: Nonlinear Phenomena*, 404:132306, 2020.
- [25] Sima Siami-Namini, Neda Tavakoli, and Akbar Siami Namin. The performance of lstm and bilstm in forecasting time series. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 3285–3292. IEEE, 2019.
- [26] Zabit Hameed and Begonya Garcia-Zapirain. Sentiment classification using a single-layered bilstm model. *Ieee Access*, 8:73992–74001, 2020.
- [27] Prajit Ramachandran, Niki Parmar, Ashish Vaswani, Irwan Bello, Anselm Levskaya, and Jon Shlens. Stand-alone self-attention in vision models. *Advances in Neural Information Processing Systems*, 32, 2019.
- [28] Alexander Alexis Suárez León and José Ricardo Núñez Alvarez. 1d convolutional neural network for detecting ventricular heartbeats. *IEEE Latin America Transactions*, 17(12):1970–1977, 2019.
- [29] Serkan Kiranyaz, Turker Ince, Osama Abdeljaber, Onur Avci, and Moncef Gabbouj. 1-d convolutional neural networks for signal processing applications. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8360–8364. IEEE, 2019.
- [30] SL Calcaterra, S Scarbro, ML Hull, AD Forber, IA Binswanger, and KL Colborn. Prediction of future chronic opioid use among hospitalized patients. *Journal of general internal medicine*, 33(6):898–905, 2018.
- [31] Ethel AM Mensah, Musarath J Rahmathullah, Pooja Kumar, Roozbeh Sadeghian, and Siamak Aram. A proactive approach to combating the opioid crisis using machine learning techniques. In *Advances in Computer Vision and Computational Biology*, pages 385–398. Springer, 2021.
- [32] Brendan Saloner and Shankar Karthikeyan. Changes in substance abuse treatment use among individuals with opioid use disorders in the united states, 2004-2013. *Jama*, 314(14):1515–1517, 2015.
- [33] Abdullah H Alzeer, Josette Jones, and Matthew J Bair. Review of factors, methods, and outcome definition in designing opioid abuse predictive models. *Pain Medicine*, 19(5):997–1009, 2018.
- [34] Sungsoo Ahn, Shell Xu Hu, Andreas Damianou, Neil D Lawrence, and Zhenwen Dai. Variational information distillation for knowledge transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9163–9171, 2019.

- [35] Geoffrey Hinton, Oriol Vinyals, Jeff Dean, et al. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2(7), 2015.
- [36] Liang Gao, Kele Xu, Huaimin Wang, and Yuxing Peng. Multi-representation knowledge distillation for audio classification. *Multimedia Tools and Applications*, 81(4):5089–5112, 2022.
- [37] Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.
- [38] Kimseth Seu, Mi-Sun Kang, and HwaMin Lee. An intelligent missing data imputation techniques: A review. *JOIV: International Journal on Informatics Visualization*, 6(1-2):278–283, 2022.
- [39] Titin Agustin Nengsih. Influence of missing data on determining the number of components for a pls regression with mcar and mar mechanism. 2019.
- [40] Young-Seob Jeong, Minjun Jeon, Joung Ha Park, Min-Chul Kim, Eunyoung Lee, Se Yoon Park, Yu-Mi Lee, Sungim Choi, Seong Yeon Park, Ki-Ho Park, et al. Machine-learning-based approach to differential diagnosis in tuberculous and viral meningitis. *Infection & Chemotherapy*, 53(1):53, 2021.
- [41] Menelaos Konstantinidis, Lisa W Le, and Xin Gao. An empirical comparative assessment of inter-rater agreement of binary outcomes and multiple raters. *Symmetry*, 14(2):262, 2022.
- [42] Juan Antonio Lossio-Ventura and Jiang Bian. An inside look at the opioid crisis over twitter. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1496–1499. IEEE, 2018.
- [43] Abeer Sarker, Graciela Gonzalez-Hernandez, and Jeanmarie Perrone. Towards automating location-specific opioid toxicsurveillance from twitter via data science methods. *Studies in health technology and informatics*, 264:333, 2019.
- [44] Sean D Young, Kai Zheng, Larry F Chu, and Keith Humphreys. Internet searches for opioids predict future emergency department heroin admissions. *Drug and alcohol dependence*, 190:166–169, 2018.
- [45] Pratap Chandra Sen, Mahimarnab Hajra, and Mitadru Ghosh. Supervised classification algorithms in machine learning: A survey and review. In *Emerging technology in modelling and graphics*, pages 99–111. Springer, 2020.

- [46] Bin Wang, Angela Wang, Fenxiao Chen, Yuncheng Wang, and C-C Jay Kuo. Evaluating word embedding models: methods and experimental results. *APSIPA transactions on signal and information processing*, 8, 2019.
- [47] Yong Yu, Xiaosheng Si, Changhua Hu, and Jianxun Zhang. A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation*, 31(7):1235–1270, 2019.
- [48] Serkan Kiranyaz, Onur Avcı, Osama Abdeljaber, Turker Ince, Moncef Gabbouj, and Daniel J. Inman. 1d convolutional neural networks and applications: A survey. *ArXiv*, abs/1905.03554, 2019.
- [49] Len Du. How much deep learning does neural style transfer really need? an ablation study. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3150–3159, 2020.
- [50] Jarek Duda. Sgd momentum optimizer with step estimation by online parabola model. *arXiv preprint arXiv:1907.07063*, 2019.
- [51] D Vijendra Babu, C Karthikeyan, Abhishek Kumar, et al. Performance analysis of cost and accuracy for whale swarm and rmsprop optimizer. In *IOP Conference Series: Materials Science and Engineering*, volume 993, page 012080. IOP Publishing, 2020.
- [52] Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. On the variance of the adaptive learning rate and beyond. *arXiv preprint arXiv:1908.03265*, 2019.
- [53] Alex Labach, Hojjat Salehinejad, and Shahrokh Valaee. Survey of dropout methods for deep neural networks. *arXiv preprint arXiv:1904.13310*, 2019.
- [54] Helena Hansen, Carole Siegel, Joseph Wanderling, and Danae DiRocco. Buprenorphine and methadone treatment for opioid dependence by income, ethnicity and race of neighborhoods in new york city. *Drug and alcohol dependence*, 164:14–21, 2016.
- [55] Marc R Larochelle, Svetla Slavova, Elisabeth D Root, Daniel J Feaster, Patrick J Ward, Sabrina C Selk, Charles Knott, Jennifer Villani, and Jeffrey H Samet. Disparities in opioid overdose death trends by race/ethnicity, 2018–2019, from the healing communities study. *American journal of public health*, 111(10):1851–1854, 2021.
- [56] Matthew A Davis, Lewei A Lin, Haiyin Liu, and Brian D Sites. Prescription opioid use among adults with mental health disorders in the united states. *The Journal of the American Board of Family Medicine*, 30(4):407–417, 2017.

- [57] Matthew A Davis, Lewei A Lin, Haiyin Liu, and Brian D Sites. Prescription opioid use among adults with mental health disorders in the united states. *The Journal of the American Board of Family Medicine*, 30(4):407–417, 2017.
- [58] Stacy Loeb, Walter Cazzaniga, David Robinson, Hans Garmo, and Pär Stattin. Opioid use after radical prostatectomy: nationwide, population based study in sweden. *The Journal of Urology*, 203(1):145–150, 2020.
- [59] Margaret R Spencer and Sierra Weathers. Trends and risk factors of adolescent opioid abuse/misuse: understanding the opioid epidemic among adolescents. *International Journal of Adolescent Medicine and Health*, 33(4), 2021.