IDENTIFY OPIOD USE PROBLEM

Abdullah Hamad Alzeer

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the School of Informatics and Computing,
Indiana University

December 2018

Accepted by the Graduate Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

Josette Jones, Ph.D., Chair

Brian Dixon, Ph.D.

September 11, 2018

Matthew Bair, MD., MS.

Xiaowen Liu, Ph.D.

DEDICATION

To those I admire the most: my parents, Hamad and Hessah, and my beloved siblings who believed in me all along, Ibraheem, Saleh, Abdulaziz, Tarfah, Ahmed, and Sarah.

ACKNOWLEDGMENTS

Everyone in this life has his journey, and the road to completing this dissertation was my odyssey. For the past 6 years, I have encountered enormous challenges that I would not have overcome without the wisdom and support of my committee members: Josette Jones, Brain Dixon, Mathew Bair and Xiaowen Liu. They have mastered the art of babysitting my ego while I sought to maintain the highest academic standards in research. Thank you very much; your lessons went beyond the pages of this dissertation, and I promise to pass the torch in the future.

In addition to my committee members, one scholar and best friend has helped the most in this journey. Lina Alfantoukh, without your support and expertise in data management, this dissertation may not have been accomplished on time. One life may not be enough to thank you.

I am grateful to all my colleagues and professors at the School of Informatics for their academic support and to the employees at the Office of International Affairs for going beyond expectations to support the international community at IUPUI.

Finally, this research would not have been possible without the trust of and financial support from the government of Saudi Arabia, represented by the College of Pharmacy at King Saud University. May I use what I've learned for the good of health research and practices in Saudi Arabia and the world.

Abdullah Hamad Alzeer

IDENTIFY OPIOID USE PROBLEM

The aim of this research is to design a new method to identify the opioid use problems (OUP) among long-term opioid therapy patients in Indiana University Health using text mining and machine learning approaches. First, a systematic review was conducted to investigate the current variables, methods, and opioid problem definitions used in the literature. We identified 75 distinct variables in 9 models that majorly used ICD codes to identify the opioid problem (OUP). The review concluded that using ICD codes alone may not be enough to determine the real size of the opioid problem and more effort is needed to adopt other methods to understand the issue. Next, we developed a text mining approach to identify OUP and compared the results with the current conventional method of identifying OUP using ICD-9 codes. Following the institutional review board and an approval from the Regenstrief Institute, structured and unstructured data of 14,298 IUH patients were collected from the Indiana Network for Patient Care. Our text mining approach identified 127 opioid cases compared to 45 cases identified by ICD codes. We concluded that the text mining approach may be used successfully to identify OUP from patients clinical notes. Moreover, we developed a machine learning approach to identify OUP by analyzing patients' clinical notes. Our model was able to classify positive OUP from clinical notes with a sensitivity of 88% on unseen data. We concluded that the machine learning approach may be used successfully to identify the opioid use problem from patients' clinical notes.

Josette Jones, Ph.D., Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

A. Synopsis

Medical and non-medical use of opioids has increased in the United States (U.S.) for the last decade for patients with chronic, non-cancer pain [1–4]. Prescriptions of opioid analgesics, such as morphine, fentanyl, oxycodone, and hydromorphone, have increased dramatically as shown in multiple studies [1,5–7]. According to The National Health and Nutrition Examination Survey, 25% of Americans 20 years and older have experienced pain that has lasted over 24 hours in the past month [8]. Despite the clinical importance of opioids in the management of pain, opioids may have significant and adverse societal effects. Notably, these can include the epidemic of opioid abuse and opioid use disorder with related socio-economic and criminal impacts, as well as deaths attributed to overdose [9]. Achieving effective pain management, along with maintaining functionality and healthy participation in society, cannot be accomplished with the high occurrence of opioid addiction. Therefore, understanding the risks and benefits of prescribing opioids is important to reduce the related opioid abuse epidemic [10–13]. The financial burden associated with the opioid use problems (OUP) is significant. The cost of treatment for patients with OUP represents a large portion of the cost of rehabilitation programs, social consequences, public safety, legal proceedings, and so on [14,15]. The total US societal costs of prescription opioid OUP were estimated to be $11.8 billion in 2001, increasing to $55.7 billion in 2009 [16]. The societal effect of prescribed opioids extends beyond financial burden to health matters, according to the National Institute of Drug Abuse; prescribed opioids were responsible for over 16,000 deaths in 2014, alone. This number increased from nearly 6,000 deaths in 2001 [17].

B. **Overview of the Opioid Use Problems (OUP)**

a. **Opioid Overdose Death**

In 2015, according to the Centers for Disease Control and Prevention (CDC), drug overdose was the leading cause of accidental death in the U.S., with 55,403 death incidents [18]. Opioid addiction alone is responsible for 20,101 (36%) of these overdose death incidents [19]. The CDC has concluded that 91 Americans die every day using prescription opioids and heroin, together [19]. To understand the relationship between heroin and OUP, according to Jones (2013), 80% (four out five) of new heroin users have started with prescribed painkillers, such as opioids [20]. Figure 1.1 shows the pattern of drug overdose deaths involving opioids by type of opioid in U.S from 2000 to 2014. [21] Unfortunately, a large part of the opioids that are available illicitly are originate from prescribed opioids. In 2012, 259 million prescriptions were prescribed for opioids [22]. This quantity is sufficient to supply every American adult with his or her own bottle [18].

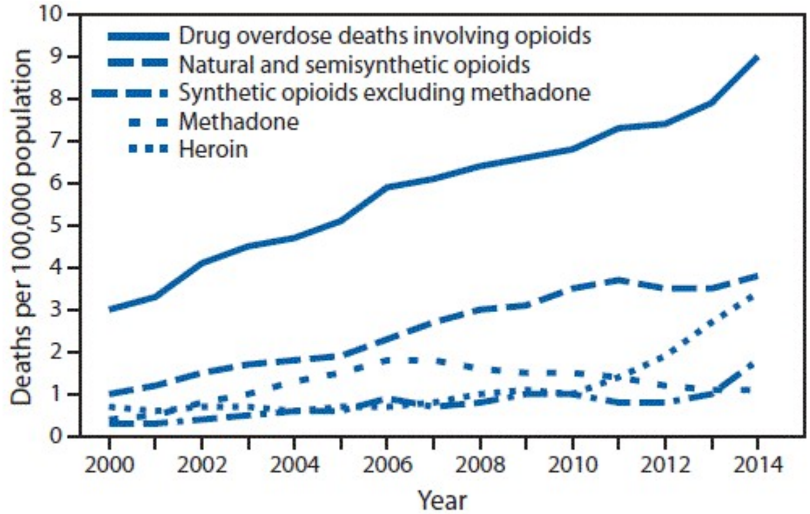


Figure 1.1 Increases in drug and opioid overdose deaths – U.S., 2000–2014

b. **Opioid Use Problems and Women**

Drug abuse affects men and women in different ways; despite men being more likely to die of painkiller overdose, between 1999 and 2010, the overdose rate among women has increased by more than 400% compared to men, whose rate has increased 256% [23]. Moreover, women may become dependent on opioids more quickly than men and engage in doctor shopping more than men, as well [23]. Another important factor unique to women is pregnancy; addicted pregnant women can put infants health at risk from neonatal abstinence syndrome [23].

c. **Opioid Use Problems in Adolescents and Young Adults**

Despite the fact that the focus of this proposal is adult chronic patients, it is important to understand that opioid consumption also affects adolescents. In 2015, an estimated 276,000 adolescents between ages 12 and 17 misused opioids, with 122,000 having an addiction to prescription pain relievers [18]. Moreover, the number nearly tripled to 829,0000 young adults from age 18 to 25 in one month (Figure 1.2) [24]. According to the National Institute of Drug Abuse, most adolescents who misuse opioids obtained their drugs for free from a friend or relative [25]. Adolescents can obtain their own prescriptions, as well. In fact, from 1999 to 2010, the number of opioid prescriptions among adolescent has nearly doubled [26].

d. **The Opioid Use Problems in Indiana**

The State of Indiana has its share of opioid epidemics; a recent report from the Richard M. Fairbanks School of Public Health (September 2016) has reported shocking facts regarding OUP. According to the source, between 1999 to 2014, the number of opioid overdoses has increased by 600% [27] (Figure 1.3). In fact, the leading cause of injury deaths in Indiana is poisoning, and most of these deaths are drug overdoses (9 out of 10) [27]. Drug overdoses overtook the number of motor vehicle deaths in Indiana [27]. Most of these drug overdoses were in the group aged 30-39, followed by the group aged 50-59 [27]. With regard to gender, more men die from drug overdoses; however, the gap has shrunk over the years [27]. Heroin deaths have also increased since 2007 (Figure 1.3); most were male, non-Hispanic whites, and people between the ages of 30-39 had the highest death rate from heroin [27]. More interesting facts can be found in the report [27].

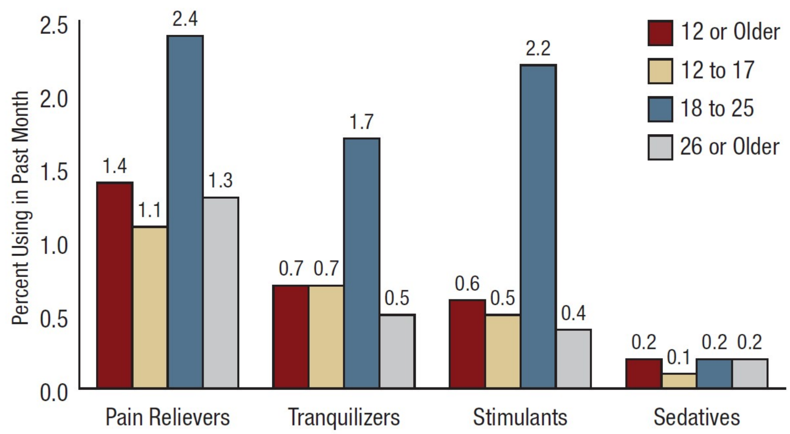


Figure 1.2 Key substance use and mental health indicators in the U.S.: Results from the 2015 national survey on drug use and health. Source: SAMHSA, 2016

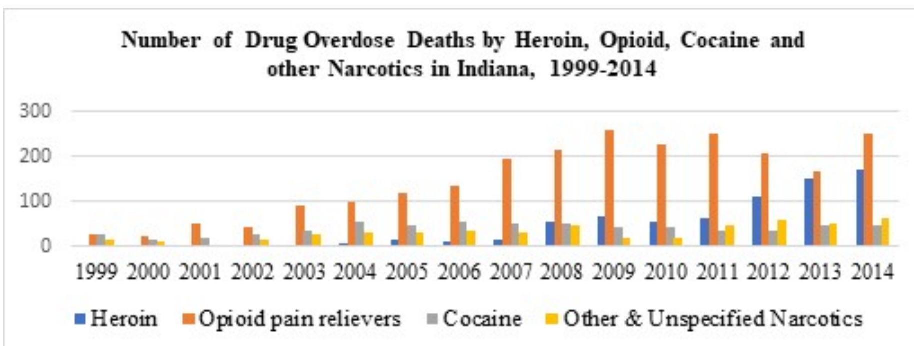


Figure 1.3 Report on the toll of opioid use in Indiana and Marion county. Source: Duwve, J. et al., 2016

C. Problem Statement

a. Physicians' Role in Identifying Opioid Use Problems

Pain-specialist physicians would be best suited to assess OUP overall and prescribe proper pain-management action. However, due to the limited number of painspecialist physicians and their accessibility to the general public, primary care physicians most often provide the majority of pain care in the health system [28]. In a study published in 2016 investigating the prescribing of schedule II medications among types of physicians, it was found that family practitioners prescribe over 15.3 million prescriptions; whereas internal medicine practitioners prescribe over 12.7

million prescriptions [29] (Figure 1.4). This prescribing may be due to improper specialized training and educational/technical support, among other factors. However, despite various root casues of the issue, primary care physicians have reported frustration when providing care for chronic-pain patients [30–34]. In a qualitative study published in 2015, some physicians admitted they have been manipulated many times and have prescribed opioids for patients who exaggerated their pain or for other patients who had tested positive several times for heroin. However, physicians sometimes make the mistake of not prescribing medication to patients who actually are in need, according to the same qualitative study [28]. Therefore, relying on clinical judgment alone might misguide physicians to choose the correct course of action to manage chronic-pain patients.

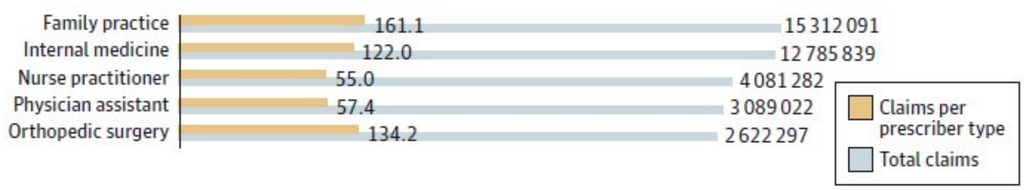


Figure 1.4 Distribution of opioids by different types of Medicare prescribers

b. **Self-Report Opioid Risk-Assessment Tools**

Multiple self-report risk-assessment tools have been designed and validated to some extent in recent years [35–38]. Existing opioid risk-assessment tools vary in the source of identifying predicting variables (knowledge base vs. database prediction) and the purpose of use (initiation of treatment vs. monitoring of care). For the purpose of this dissertation, we categorize opioid risk-assessment tools into two categories: self-report base vs. database tools. We define self-report-based risk-assessment tools as risk-assessment tools that predict the risk of current or future opioid use problems based on predictive variables extracted from data self-reported by the patient or physician. We define database risk-assessment tools as risk-assessment tools that predict the risk of current or future opioid use problems based on objective data extracted from a database, such as electronic health records or claims administrative data (Figure 1.5).

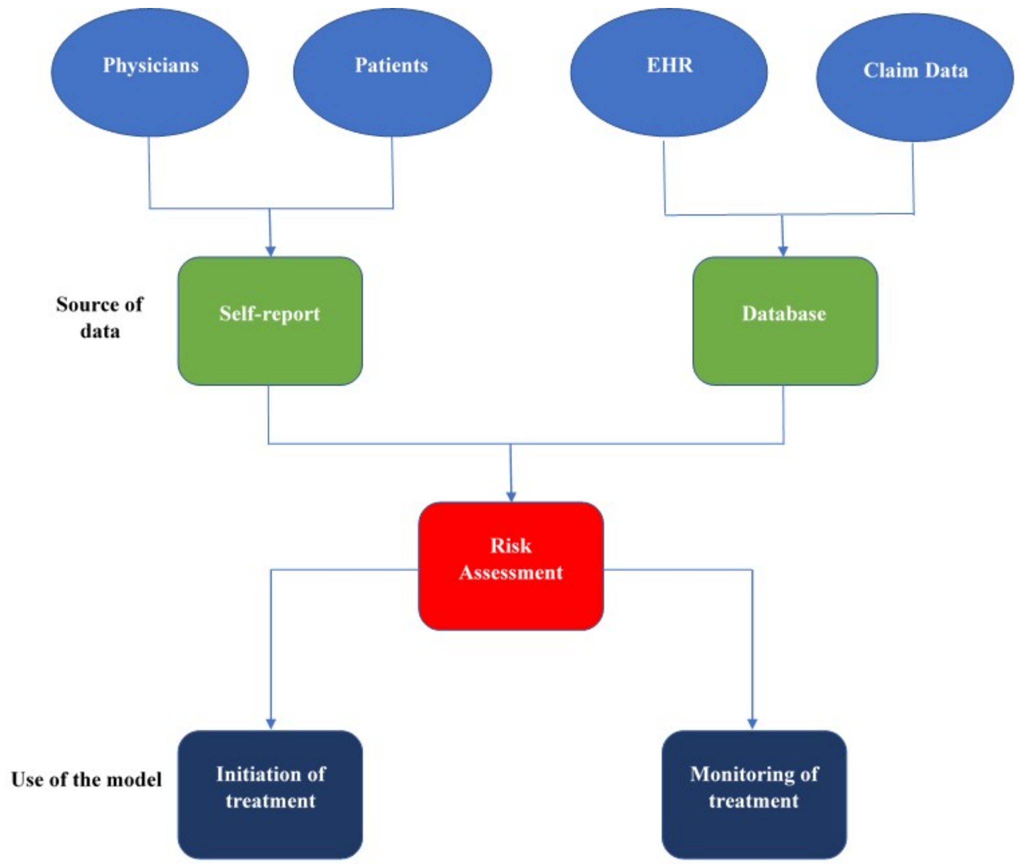


Figure 1.5 Overview of sources and purpose of opioid risk assessment tools

Common examples of self-report tools are SOAPP, ORT, and COMM. However, having the majority of risk-assessment tools solely rely on patient completion might limit their ability to predict aberrant drug behavior due to reporting bias [39] or lack of patient comprehension [40]. Jones et al., (2012) conducted a study that is unique in that it compared original ORT (patient self-reported) vs. ORT (psychologist completed after conducting an interview) among 51 patients; only 30 (59%) patients matched in the same risk category. Among all variables, age had the highest rate of agreement, with only an 86% level of agreement [40]. Overall predictive ability varied, as well between the two ORTs; predictive ability for psychologist-completed ORT was better (43% vs. 70% missing rate) [40]. These results indicate that the process of administration of the opioid risk-assessment tool itself could play a role in determining the overall predictive ability of the tool; perhaps patient education and leveraging physicians with more accurate and robust data from health care databases might help to minimize the issue. A second issue of using self-report methods to assess opioid risk is the process of documentation itself. This challenge becomes a monthly burden on the clinical staff, nurses, and physicians who are conducting the documentation process. Understanding the challenge of documentation, Butler conducted a study to examine the effect of computerizing current methods of administrating SOAPP and COMM (self-report type) on the rate of documentation of these tools. The author designed the web-based tool and implemented it in two pain centers for a period of 10 months. Both sites used paper-based versions of SOAPP-R and COMM. The results showed a significant increase in the documentation of opioid risk assessment for both initial and follow-up visits (SOAPP 30.3% vs. 76.9/5, COMM 4.5% vs. 43.6%) ($p < .001$). Such a study highlights the significant issue with documentation in current self-report methods [41].

A third issue with self-reported methods is the variation of prediction ability (reliability) in the literature; SOAPP is one of the most validated tools we found in the literature. Based on this review, there is a large variation in the sensitivity of the tool to predict OUP. The sensitivity and specificity of SOAPP for one source was 0.91 and 0.69 [42] and 0.392 and 0.693 in another [43], respectively. The same argument can be applied to ORT, in which one study reported a sensitivity of 0.45 [44] and another reported a sensitivity of 0.195 [43]. One way to explain this variation could be the difference in defining the outcome in the validation studies reviewed.

D. **The Statement of Purpose**

Due to the limited number of pain specialists, primary care physicians prescribe a significant portion of opioids. Primary care physicians have reported frustration when providing care for chronic-pain patients. Our objective is to inform the clinical and research community about factors contributing to the incidence of opioid use problems.

Given the lack of census on structured variables' rule to identify opioid use problems in the literature, we seek to identify studied variables in the literature and classify their rules in terms of protection or being risk factors to opioid use problems. Additionally, while 80% of health data are stored as unstructured data, the current tools to identify opioid use problems majorly rely on patient-reported data or structured data in the electronic medical records. Thus, we seek to utilize unstructured data to develop a predictive model for those who are at risk of developing opioid use problems using text mining and machine learning approaches.

E. **Dissertation Aims**

1. Conduct a systematic review of previous opioid use problems (OUP) predictive models to:
   (a) Identify variables available in the literature to predict OUP.
   (b) Explore and compare methods (population, database, and analysis) used to develop statistical models that predict OUP.

(c) Understand how outcomes were defined in each statistical model OUP.

2. Identify and compare patients with opioid use problems (OUP) using the text mining vs. ICD codes.
   (a) Identify positive opioid use problems patients using text-mining approach.
   (b) Identify positive opioid use problems patients using the ICD-9 codes.
   (c) Compare the results and population characteristics.
3. Develop a predictive model for those who are at risk of developing OUP using a machine learning approach.
   (a) Identify a gold standard subset of medical reports.
   (b) Identify best configurations to build machine learning model.
   (c) Build a model that predicts opioid use problems based on patients' clinical notes.

F. **Definition of Terms**

To understand the complexity of the problem, there are several terms related to opioid use that need to be defined: opioid addiction, opioid abuse, aberrant drug taking behavior (ADTB), and opioid use problems (OUP). According to R. West and J. Brown in their book, Theory of Addiction, 2012, addiction is defined as a chronic condition in which there is a repeated powerful motivation to engage in a rewarding behavior, acquired as a result of engaging in that behavior, that has significant potential for unintended harm. It is not all-or-none, but a matter of degree. [45].

The second term is opioid abuse. In general, substance abuse is an initial step towards addiction and dependence. The World Health Organization (WHO) defines substance abuse as the harmful or hazardous use of psychoactive substances, including alcohol and illicit drugs. [46] Some attributes that characterize substance abuse are: failure to fulfill social or work obligations, continued use of a substance in hazardous situations, legal problems related to substance abuse, and persistent use despite continued and recurrent problems [47].

Substance abuse and dependence are now combined into substance use disorder (SUD), which is measured on a continuum from mild to severe according to Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM 5). As this area of study has developed, so have the defining criteria related to substance use disorders. Categorization within the spectrum of SUD requires at least two to three symptoms - instead of one- from a list of 11 criteria for mild stage [48]. While these criteria are available in detail at www.DSM5.org [48], it is important to note that the tolerance criterion and withdrawal symptoms for opioid use disorder "does not apply for diminished effect when used appropriately under medical supervision" [49].

Aberrant drug taking behaviors (ADTB) are defined as "behaviors that are more likely to be associated with medication abuse and /or addiction" [50]. Examples of these behaviors are: 1) selling prescription drugs, 2) concurrently abusing alcohol or other illicit drugs, and 3) losing prescribed medication on multiple occasions [51–53]. It is worth mentioning that some behaviors could look like ADTB, but in fact, they might indicate undertreatment or may be more part of stabilizing the pain condition [54]. Examples of these behaviors are: 1) asking for, or even demanding, more medication, 2) asking for specific medications, 3) use of the pain medication to treat other symptoms [54]. Despite the evolving nature of these terminologies, it is important for researchers and treating prescribers to understand that there is a spectrum of addiction.

For this study, we define the final term, opioid use problems (OUP) as the presence of opioid aberrant behaviors, opioid misuse, opioid abuse and opioid use disorder in the patient electronic health record (clinical notes and/or ICD codes). A list of ICD codes and a list of specific criteria of aberrant behaviors (for the clinical note) are provided in the method section (Appendix A.3 and Table 3.1).

2. CONDUCT A SYSTEMATIC REVIEW OF PREVIOUS OPIOID USE PROBLEMS (OUP)

A. Objective

Several opioid risk assessment tools are available to prescribers to evaluate opioid analgesic abuse among patients with chronic pain. The aim of this systematic review is to answer the following questions: 1) What variables have been examined in the literature to predict opioid abuse? 2) What are the methods (population, database, and analysis) used to develop the statistical models to create the tool? 3) How were the outcomes defined in each statistical model? The results of this chapter was published in a paper titled: "Review of factors, methods, and outcome definition in designing opioid abuse predictive models. Pain Medicine, 19(5), 997-1009."

B. Method

A pharmacist with research experience used EBSCO and PubMed to conduct a two-stage systematic search to identify several articles that are directly related to the topic of risk assessment. First, the pharmacist identified 8 articles that were deemed to be directly related to the topic. The second step was to conduct a systematic search using OVID to generate a list of articles that should have all 8 articles of relevance. The idea behind choosing this method is ensuring all common articles of relevance are included in one combined search, which perhaps will yield more new articles related to the topic, as well. Finally, authors of some of the articles found were also contacted mainly through email or over the phone to require further data about their work or to identify other related work. The search was limited to articles written in English featuring human adult subjects published from January 1990 to April 2016.

This search generated 1409 articles. Duplicate articles were automatically checked using Endnote X7 and then manually reviewed. Out of 1409 articles, 43 articles were duplicates and excluded. The pharmacist reviewed the titles and abstracts of the remaining 1366 articles (Figure 2.1). The inclusion/exclusion criteria were the following:

1. The included articles were original studies, not narrative reviews.
2. The included studies focused on risk for aberrant drug-related behaviors.
3. Studies included quantitative data specific to prediction capability (odds ratio, p-value, or confidence interval).
4. Studies that focused on self-administered tools were excluded.
5. Only studies that used algorithms to predict opioid abuse from data extracted from electronic health records or administrative claims were included.

After applying the inclusion/exclusion criteria, 27 full-text articles were downloaded. Of these articles, 20 articles were excluded, because of their focus on selfreport tools (Figure 2.1). Variables of interest were extracted and listed on an Excel spreadsheet for all 7 articles included. To check face validity and data consistency, a primary care physician reviewed extracted data, including variables from articles, categorization of variables from articles, method (population, database, analysis) used, and outcome definition (opioid abuse).

C. **Results**

During our systematic review, we identified seven articles [55–61] (nine models) to assess opioid risk from databases (claims data or EHR), which will be the subject of this review. All nine models provided definitions of the outcome of opioid abuse as well as variables used to predict the outcome. The following sections will discuss important findings with regard to variables, methods, and outcome definitions for the coinciding models.

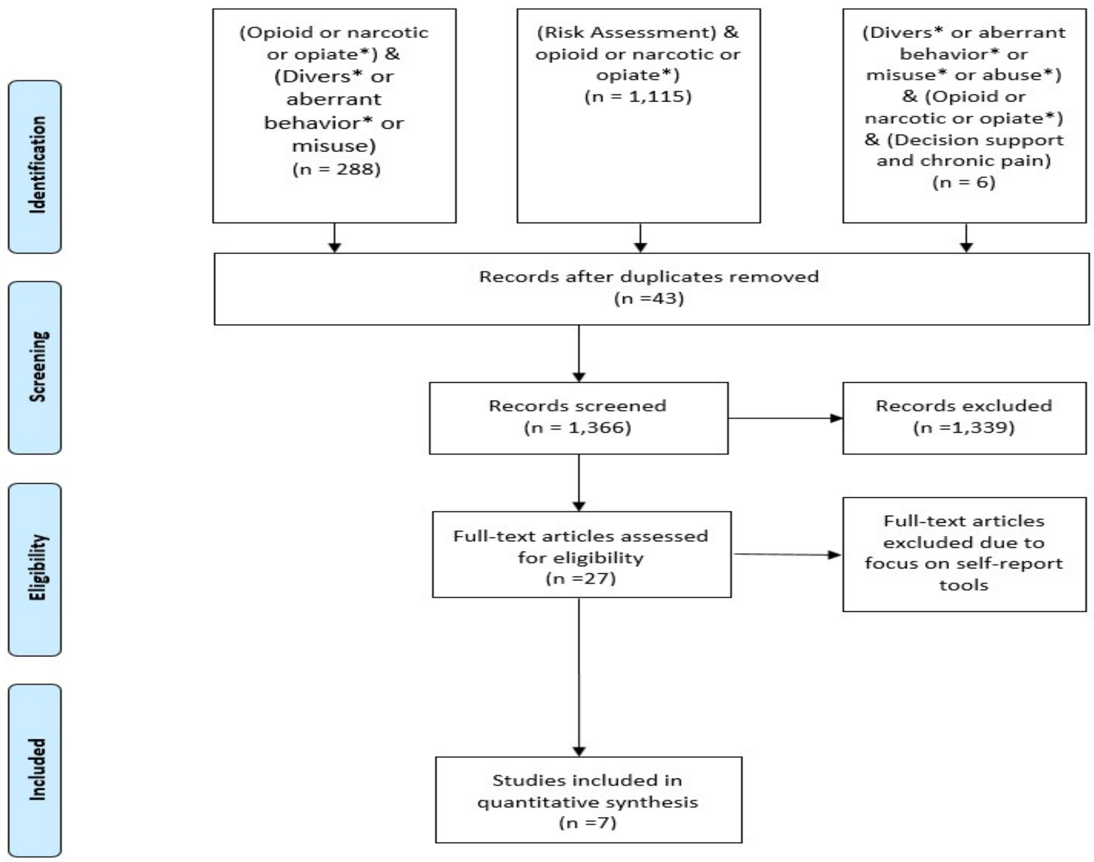


Figure 2.1 Articles extraction process

a. **Variables**

Among the retrieved studies, 9 models and 75 distinct variables were identified. The variables were grouped into 7 main categories; demographics (6 variables), medications (33 variables), care utilization visits (3 variables), behavior (7 variables), mental status (1 variable), pain and medical comorbidities (15 variables), pain (4 variables), and family history of substance abuse and comorbidities (6 variables).

The identified models most commonly included demographic variables; age and gender were mentioned in all 9 models. History of alcohol abuse, smoking status and mental diagnosis were mentioned in 5 models. Table 2.1 lists all variables mentioned in 3 models or more out of 9 models.

b. **Method: Studied Population, Database Used and, Analysis**

As mentioned previously, this review found 3 studies that used administrative claims data and 4 studies that used electronic health record data. In 1 of the 4 articles, a disease management program database was used in addition to health record data. The number of patients included in the studies varied greatly from 196 to 1,552,489. Studies that used ICD-9 codes as a primary source of defining their cohort included larger sample sizes compared to studies that did not use ICD codes. This sample size variation is related to study design and coincides with the nature of data used. Administrative claims data is likely national data, while EMR data is representative of a single site or single health care system. The inclusion criteria for a study's population were similar, age from 12 (or 18) to 64, at least one claim for a prescription opioid in the past 12 months, or a history of receiving opioids for 30 to 90 consecutive days. Exclusion criteria were cancer pain and heroin poisoning. Table 2.2 details the methods for each article. Finally, in terms of analysis used, all studies used logistic regression as their primary multivariate analysis method. Using logistic regression for this purpose is consistent with the common practice in the field, which recommends using logistic regression to analyze a dichotomous dependent variable [62] (e.g. opioid abusers vs. non-abusers). Some of the studies preliminarily analyzed variables in bivariate analysis, with the significant variables ultimately entered into the multivariate model.

Table 2.1 Most commonly used variables in statistical models to predict opioid abuse

Category	Variable	Interpretation (Risk = Opioid abuse)	# of models	Times it was protective	Times it was risk factor	Odds not reported or = exactly 1
Demographics	Age	Older age decreases risk	9	9	0	0
Demographics	Gender	Male increases risk	9	0	6	3
Behavior	History of ethanol abuse	Alcohol abuse increases risk	5	1	3	1
Behavior	Smoking status	Smoking increases risk	5	0	4	1
Mental	Mental health diagnosis	Mental disorder increases risk	5	0	4	1
Medications	# of opioid prescriptions	Increase number of opioid prescription increases risk	4	0	4	0
Behavior	Early refills of opioid prescriptions	Increase number of early refills of opioid prescription increases risk	4	0	4	0
Medications	Non-opioid substance abuse/dependence	Non-opioid substance abuse/dependence increases risk	4	0	4	0
Care utilization	Inpatient hospitalization days	Greater hospitalization days increases risk	3	0	3	0
Care utilization	Days with physical care visits	Greater outpatient visits increases risk	3	0	3	0
Behavior	# of pharmacies where opioid prescription were filled	Greater number of pharmacies where opioid prescription were filled increases risk	3	0	3	0
Behavior	# of opioid prescribers	Greater number of opioid prescription increases risk	3	0	3	0
Medications	Methadone	Presence of methadone prescription increases risk	3	1	2	0
Medications	Fentanyl (IR)	Presence of fentanyl prescription has no effect on risk	3	1	1	1
Medications	Fentanyl (ER)	Presence of Fentanyl (ER) prescription increases risk	3	0	2	1

Category	Variable	Interpretation (Risk = Opioid abuse)	# of models	Times it was protective	Times it was risk factor	Odds not reported or = exactly 1
Medications	Morphine (IR)	Presence of morphine prescription increases risk	3	0	2	1
Medications	Morphine (ER)	Presence of Morphine (ER) prescription increases risk	3	0	2	1
Medications	Hydromorphone	Presence of Hydromorphone prescription has no effect on risk	3	1	1	1
Medications	Oxycodone	Presence of Oxycodone prescription decreases risk	3	1	0	2
Medications	Tramadol	Presence of Tramadol prescription decreases risk	3	3	0	0
Medications	Codeine	Presence of Codeine prescription decreases risk	3	2	0	1
Medications	Propoxyphene	Presence of Propoxyphene prescription decreases risk	3	2	0	1
Medications	Hydrocodone	Presence of Hydrocodone prescription decreases risk	3	2	0	1
Medications	Morphine (ER)	Presence of Morphine (ER) prescription increases risk	3	0	2	1
Medications	Fentanyl (ER)	Presence of Fentanyl (ER) prescription increases risk	3	0	2	1
Medical Comorbidities	Hepatitis	Presence of Hepatitis diagnosis increases risk	3	0	3	0

c. **Outcome Definition**

The examined articles defined the outcome of opioid abuse differently. Specifically, 4 out of 7 articles (6 out of 9 models) depended primarily the on presence or absence of an opioid abuse or dependence diagnosis according to ICD-9 codes, while the other 2 studies used a predefined list of opioid-related aberrant behaviors. Additionally, they also reviewed a patient's profile for any release from the pain-management program due to aberrant behaviors. The last article and respective model, used a hybrid technique that included natural language processing methods, along with ICD-9 codes to define opioid problems (Table 2.3). [55–61]

Table 2.2 Outcome definition and methods data source

Article/ # of Models	# of Patients	Study sample	Source of data/Period	Coding
Lewis, 2014/1	202	Prescriptions of an opioid for 30 or more days of consecutive dosing. Age greater than 18 years, and a documented history of any chronic pain syndrome defined as pain lasting greater than 90 days in duration.	Electronic Health Record/12 Months	Manual Chart review
Ives, 2006/1	196	chronic, non-cancer pain of at least three months duration	Medical record and our disease management program database/12 months	Followup
White, 2009/2	632,000	12 to 64 years with at least 1 claim for a prescription opioid and at least 1 medical claim from 2005 to 2006.	Administrative claims data/3 months (Alternative A) & 12 months (Alternative B)	ICD-9- CM
White, 2012/2	1,552,489	Patients ages 12-64 years throughout the 12 months prior to the index opioid Rx	Administrative claims data/ 12 Months	ICD-9- CM
Edlund, 2007/1	15,160	Veterans with at least one prescription for an opioid and had more than 91 days of supply in 2002. (excluded individuals with any cancer diagnosis (ICD-9-CM codes between 140.0 and 208.9)	Administrative claims data/ 24 months	ICD-9- CM
Turner, 2014/1	5,420 Urine Drug tests	Patient Age 20 or above. Study included only Urine Drug Test (UDTs) for patients on Chronic opioid therapy (COT) defined by Group Health as 70 days opioid supply in the prior 90 day. To limit the sample to UDTs for CNCP patients, the study excluded those of patients who, in the one-year period prior to the UDT, had had hospice care, opioid prescriptions from oncologists, or more than one visit for cancer other than non-melanoma skin cancer.	Electronic Health Record/ 12 months	ICD-9
Hylan, 2015	2,752	18 years or older who initiated COT for noncancer pain between 2008 & 2010. COT was defined as receipt of equal or more than 70 days supply of transdermal or oral opioids (except buprenorphine) in a calendar quarter, which corresponded to >75% of the days in the quarter covered by an opioid prescription. Study patients receive at least 2 quarters of COT within a 1-year period	ICD-9 codes: 304 [0, .00, .01, .02; 304.7, .70, .71, .72]; 305 [5, .50, .51, .52]	ICD-9 + Natural Language Processing (NLP) to identify [overuse, misuse, abuse or addiction] And/ or [Dependance]

Table 2.3 Outcome definition and methods data source

Article /# of Models	**Outcome Definition**
Lewis, 2014/1	multiple pharmacies, multiple non Belleville Family Health Center (BFHC) prescribers of opioids (any prescription preceded and followed by a BFHC provider prescribed opioid), unsanctioned dose escalation, lost or stolen prescription, UDS anomalies, illicit drug use (either self-reported or identified via UDS), multiple emergency department visits resulting in receipt of opioids, family report of opioid sale or other form of diversion, Forging or tampering with controlled substance, prescriptions and misuse/adulteration of prescribed formulation or intended route of administration.
Ives, 2006/1	Negative urine toxicological screen (UTS) for prescribed opioids, UTS positive for opioids or controlled substances not prescribed by our practice, Evidence of procurement of opioids from multiple providers, Diversion of opioids Prescription forgery , Stimulants (cocaine or amphetamines) on UTS.
White, 2009/2	304.0 (opioid-type dependence), 304.7 (combinations of opioid type with any other), 305.5 (opioid abuse), or 965.0 (poisoning by opiates or related narcotics but excluding 965.01 [heroin poisoning])
White, 2012/2	ICD 9 codes: 304.0, 304.7, 305.5, 965.00, 965.02, and 965.09
Edlund, 2007/1	ICD 9 codes: 304.00304.03 (opioid dependence), 304.70304.73 (dependence; combinations of opioid type drug with any other), and 305.50305.53 (opioid abuse)
Turner, 2014/1	Clinical Classification Software (CCS) categorization. Smoking Status used 305.1 ICD-9 code.
Hylan, 2015	ICD-9 codes: 304 [0, .00, .01, .02; 304.7, .70, .71, .72]; 305 [5, .50, .51, .52]

D. **Discussion**

a. **Inconsistency of Prediction Direction of Variables**

Of note, we found inconsistency in the ability of individual variables to predict the outcome (aberrant drug-related behavior). Some variables, such as using methadone (for opioid use disorder treatment) were significant risk factors in two models (OR= 2.97, CI= 2.57-3.42, p-value < 0.0001) and (OR= 3.22, CI= 2.62-3.95, p-value < 0.0001), and significantly protective factor in a third model (OR= 0.26, CI= 0.08-0.94, p-value = 0.02). A second example is seen with the variable history of alcohol abuse. Despite a predominance in the literature, which shows that a history of alcohol abuse is a risk factor (4 studies), the same variable was slightly protective in a fifth model. This variation in the role of a predictive variable (protective vs. risk factor) might be solved by having a wider measurement scale (continuous or multiple categories) for the variables instead of a dichotomous scale (Yes or No). To support this argument, Zale, E. et. al., 2014, conducted the first study of its kind to test whether varying levels of current/historical smoking and indices of smoking heaviness/nicotine dependence may be associated with greater likelihood of past-year prescription opioid misuse in the general population. [63] The study found a positive association between level of smoking heaviness/nicotine dependence and opioid misuse [63]. Other variables, such as gender, number of opioid prescriptions, early refills, inpatient hospitalization days, days with physical care visits, non-opioid substance abuse/dependence, number of morphine prescriptions, and number of opioid prescribers were all consistent risk factors across the models. Other variables, such as age (older groups) and tramadol use, were consistently reported as protective variables (Table 2.1).

b. **Claims Data vs. Electronic Health Record Data**

Using a specific type of database (administrative claims data vs. electronic health record data) could potentially alter development of the model design. For example, administrative claim databases generally provide a consistent data format, because they use a pre-defined set of codes, [64] As result, it is relatively easier to create a study cohort using a claims data compared to most electronic health records databases. However, one potential disadvantage of using claims data is that the care providers do not always document these relevant codes. This incomplete documentation leads to missing outcomes of interest for some patients. [65,66] On the other hand, using electronic health record databases, which has much richer clinical data (for each patient), [64] can facilitate improved cohort/outcome definitions. Specificity within the cohort can be achieved by reviewing some aspects of the patient's file manually or using some advanced tools to define the study cohort based on certain metrics (e.g., estimated glomerular filtration rate to define chronic kidney failure patients) [64].

c. **Variation in Outcome Definition**

The International Classification of Diseases (ICD) is a diagnostic coding system that can be used to categorize patients. However, the issue with creating a model based on an outcome structured by ICD codes is the fact that these codes often understate the actual number of patients exhibiting the target categorization [57,58].

This means those patients who are labeled with opioid problems, according to ICD codes, are likely to have an opioid problem; however, ICD codes may overlook other patients who have opioid problems but were not documented [57]. This could be due an intentional lack of documentation, or in other cases, prescribers try to avoid stigmatizing patients with opioid problems using these codes. [57] The issue of underdocumentation in particular can create less representative models. Another limitation to the models found is the fact that most of these models were based on claims data, which is less likely to provide the in-depth and detailed clinical data available in an electronic health record database. The variation in the outcome definition across studies and its potential effect can be noticed in the self-report opioid risk assessment tools, as well. When tools, such as SOAPP-R (SOAPP revised) or ORT are validated across different populations or when different definitions of the outcome are used, the prediction sensitivity and specificity changes greatly. SOAPP-R is a self-report questionnaire designed to predict aberrant medication-related behaviors among persons with chronic pain [67]. SOAPP-R came as a revised version for the original SOAPP, an instrument developed by Inflexxion, (Newton, MA, USA) with support from Endo Pharmaceuticals and the National Institute on Drug Abuse [68]. Version 1, considered as an initial step toward development of a screener for aberrant medication-related behaviors in chronic pain patients [37]. SOAPP-R is one of the most validated tools; based on literature, there is a large variation in its sensitivity. The sensitivity and specificity of SOAPP were 91% and 69%, respectively, in one study [42], and 39.2% and 69.3%, respectively, in another [43]. The same statement can be applied to ORT, as well. One study reported sensitivity of 45% [46]

and the other study reported sensitivity as low as 19.5% [43]. One way to explain the variation in opioid risk assessment tools performance could be due to a difference in outcome definitions in the reviewed validation studies. For example, SOAPP performance was measured in some studies based on a positive result on the Aberrant Drug Behavior Index (ADBI) [67–69]; while in other studies, it was measured based on the presence of a discharge review form [70]. At present, such cross-comparisons for opioid risk assessment models, which are based on databases, are not yet possible due to a lack of further validation from multiple studies for any of the models found. However, this review indicates that more effort in the future is likely to utilize secondary-data analysis. This is due to the necessity of developing more accurate and comprehensive models from current and future health system electronic records.

However, these future efforts will surely face an interesting challenge, considering that, as of 2016, most of the organizational data are unstructured, and health care data is no exception. According to IBM, in 2015, unstructured data represented 80% of the total health data. [71] Thus, it is important for future efforts to utilize the untapped potential of unstructured data to develop an opioid addiction model that uses data mining techniques.

Another aspect of defining outcomes is the cohort follow-up period. Whether the studies retrieved were retrospective or prospective, the covered period to measure the outcome ranged from 3 months in one model, including 12 months in 6 models, up to 24 months in one model, and finally 24-60 months (a combination of a pre-indexing period and post-indexing period). This variation in the period of follow up might have an unforeseeable impact on the prediction accuracy of the models. For example, a period of 3 months of follow-up may not be sufficient to detect opioid-related aberrant behaviors. Thus, patients might be categorized false negative as a result.

d. **Limitations**

There are only 7 studies, including 9 models, that met our inclusion criteria. This could be due to using only one search engine, OVID Medline. However, the study's search terms were exhaustive and used results from combined 3 search sessions to mitigate this shortcoming (Figure 2.1). Another reason for the relatively small number of articles is the exclusion of self-report opioid risk assessment methods (21 articles). The justification for omitting these articles is specific to the study questions. The study attempts to identify variables and resources used for the design of models based on databases, electronic health records, and administrative claim data. Another limitation related to data collection is having only one primary care physician to check for face validity and data consistency. Having another reviewer and reporting interrater reliability could add value to the study. In our case, this was not feasible, due to limited funds and resources available for the study.

A second limitation pertains to the quantitative analysis of the variables found, which did not include magnitude of the odd ratio or significance of p-values for each variable. Rather, the analysis only describes a variables rule in terms of being a risk factor or protective measure (Table 2.1). This was due to a lack of reporting odds ratios in 2 articles and p-value data in a third article. These 2 articles did not report the odds data for the insignificant variables, while the third article did not report any p-values for the model (but reported adjusted OR and CI, instead). Finally, the way articles categorized the variables subset varied across the model. For example, the variable age was sub-categorized into 6 age groups, while it was treated as a continuous variable in the other. Moreover, determining the reference group varied among models, as well. In the same example of the variable “Age”, one model identified an older age sub-group as a reference group, while 5 models identified a younger age sub-group as a reference, and one article identified middle age (40-49) as a reference group.

A third limitation of this review is the inconsistency in the definition of the outcome, opioid abuse, across the articles. Despite a description and analysis of the same issues, the particular definitions employed varied across articles. Shannon M. Smith et al examined the issue of definition inconsistency in 2013. [72] The study summarized differences between multiple opioid-use-disorder terminologies based on experts’ opinion. The study found that the term “misuse” “emphasizes the use of the substance does not follow medical indications or prescribed dosing” [72]. However, the term “abuse” is commonly applied to substance use for nontherapeutic purposes. Addiction, on the other hand, was defined as “compulsive substance use that occurs despite personal harm or negative consequences” [72]. To address this variation of the definition in the examined articles, we summarized outcome definitions by each article in Table 2.3. Our analysis was limited to the data provided in the articles. Thus, access to certain data that measured the models performance, such as c-statistics, sensitivity, and specificity for analyzed models, were not available.

E. **Conclusion**

Opioid risk assessment tools are becoming standard practice in pain and primary clinics prior and during prescribing opioid therapy for chronic pain. However, this review indicates that there is inter/intra variation in these tools' performance for assessing opioid abuse. This review concludes that this variation can be explained by the variation in study period, sample size, opioid abuse definition, type of database, and structured documentation to ICD codes. In addition, this review illustrates an overview of common methods of opioid risk assessment tools and categorizes them based on method of reporting into self-report and database-related models. Moreover, the study provides a comparison Between the two categories whenever possible. This systematic review presents a list of variables that were used to predict opioid abuse from electronic health records and administrative data. Furthermore, the review provided a count of how many times each variable was mentioned and how often it was counted as a risk factor or protective measure in the literature. To our knowledge, this is the only systematic review that has done so. Providing such data to researchers could lead to developing a tool that is more accurate in predicting the risk of developing aberrant drug-related behaviors. The review also identifies and compares other aspects related to opioid risk assessment models design, such as period of the study, number of patients, subject inclusion criteria, and ICD-9 code used to identify patients from the database. The study highlights major differences between articles defining opioid abuse derived from databases for the purpose of developing opioid use risk assessment tools. This could help future researchers build on previous work to create advanced models for improved predictability of opioid abuse. Despite the availability and presence of many self-report and database-oriented risk assessment tools, prescribers might not yet rely solely on these tools due to their lack of validation and consistency in their results. However, opioid risk assessment procedures can be improved by enhancing structured data captured by the electronic health record system. Physicians and nurses can play a major role in this step by documenting the proper ICD code that specifically identifies the category of opioid abuse for patients. Clinical practices and hospitals should pay attention to the viability of adopting these tools into their practice. Some expected barriers to adoption of these tools are: the population variation from within the clinic where the tool was validated, procedural differences, differences in outcome definition between clinics, and lack of proper technical experience to implement these tools.

3. IDENTIFY AND COMPARE PATIENTS WITH OPIOID USE PROBLEM

A. Introduction

Opioids are a group of medications prescribed to relieve pain. Before the 1980s these drugs were mainly prescribed by surgeons to relieve immediate postoperative pain or pain related to cancer [73,74]. However, during the 1980s, the move toward aggressive pain treatment and prescribing opioids for chronic non-cancer pain has been encouraged [73,75]. Advocacy health groups, such as American Pain Foundation, continued to push for more opioid prescriptions during the late 1990s and through the 2000s [73,76–78]. By 2012, opioid prescriptions increased from 142 million in 1999 to 248 million prescriptions [79,80] and opioid sales quadrupled from 1999 to 2010 [81]. Particular medications, such as hydrocodone, doubled in consumption between 1999 and 2011 [82].

In conjunction with the increase in prescribed opioids per population, there was also a higher percentage of opioid overdose deaths and substance use disorder treatment increased in parallel from 1999 to 2008 [19]. In 2015, drug overdose incidents accounted for 52,404 U.S. deaths, of which opioids were involved in 33,091 overdose incidents (63.1%) [83]. The 2016 National Survey on Drug Use and Health indicated that there was approximately 11.8 million people in the U.S. age 12 years or older who misused opioids in the past year [24]. At the state level, Indiana state reported 794 opioid-related overdose deaths with a substantial increase in heroin-related overdose deaths. Compared to 2012, heroin-related overdose deaths increased from 114 to 297 in 2016. Deaths from synthetic opioids have increased for the same period from 43 to 304 deaths in Indiana [84].

Despite the importance of the opioid issue, reporting prevalence of opioid problems may still impose some challenges. For example, differences in the targeted outcome definitions (“opioid use disorder” vs. “opioid overdose” vs.“opioid abuse”) and population inclusion criteria (minimum medication day’s supply or number of opioid prescriptions) may alter the study’s ability to identify the issue [85]. Despite these differences, International Classification of Disease (ICD) identification may be a standard method that may used to report opioid use disorder, overdose, abuse, and dependence (which we will refer to in this study as Opioid Use Problem (OUP)). ICD is a code system used by physicians to classify and diagnose disease. However, the issue with reporting an outcome defined by ICD codes is the fact that these codes often understate the actual number of patients exhibiting the target categorization [57,58].

Thus, we examined the literature to identify alternative methods to identify opioid use problem in addition to ICD codes. We found a limited number of studies that discuss alternative methods using natural language processing (NLP), or text mining, in a general healthcare setting [86–89]. The use of text mining in these studies varied from accelerating the documentation process of patient records to automatically parsing and coding clinical events or assisting decision-making processes by classifying specific complex diagnoses.

More relevant studies addressing NLP/text mining to identify opioid use problems were identified as well [61,90,91]. Carrell et al. (2015) developed an NLP process to identify evidence of opioid use problems in electronic health records (EHRs) at Group Health, a healthcare system in Seattle, WA, USA. The study found that conventional diagnostic codes for opioid use problems has identified 2,240 (10.1%) patients and NLP identified an additional 728 (3.1%) patients by analyzing patients' clinical notes [90]. Hylan et al. (2015) has also used NLP to "report on a predictive model developed to assess the likelihood of problem opioid use over a 2- to 5-year period following initiation of chronic opioid therapy" within the Group Health system. The study found the sensitivity of the regression model prediction of problem opioid use to be 58.3%, with specificity being 71.2% [61]. Despite these early efforts of to use NLP/text mining encouraging, no study has attempted to identify opioid use problems using text mining in a multi-site study setting.

In this study, we compared different strategies for identifying OUP using administrative and EHR data. We examined whether adding a text mining approach could improve identification of OUP for patients on long-term opioid therapy (LTOT) for a large population across a multi-site healthcare system.

B. **Method**

a. **Methods Overview**

Following Institution Review Board (IRB#: 1710876219) and an approval from the Regenstrief Institute (RI) Data Management Committee, we queried identified the patients' clinical notes from Indiana University Health (IUH), a large healthcare network in Indiana with 3,541 staffed beds and 2,563,086 outpatient visits [92]. The IUH system produces over 30 report types (e.g., Visit notes, Progress notes, Discharge notes). Next, a text-mining algorithm was applied to identify opioid use problems using patients' clinical notes and reported incidence results (Figure 3.1). Finally, to investigate whether or not there was a difference between positive OUP using a text mining approach vs. an ICD conventional approach, we compared the 2 positive cohorts characteristics in terms of frequency and rate per 1,000 long-term opioid therapy patients and highlighted key points. Additionally, we compared results of care utilization (outpatient visits, emergency department visits, hospitalizations, cumulative hospitalization days) and they were also reported.

b. **Sample Definition**

Our sample was defined as adult IUH members age 18 years or older who have been prescribed LTOT, which is defined as patients with 70 days of supply within any given 90-day period between 1 January 2013 and 31 December 2014 (24 months). We excluded patients with active cancer to focus on LTOT for non-cancer chronic pain (ICD-9 codes 140.x 172.x, 174.x 209.xx, 235.x 239.xx, 338.3). Patients with schizophrenia were also excluded due to the documented high percentage of opioid dependence among this population (ICD-9 code 295.9) [93].

c. **Variables of Interest**

Patient characteristics, including demographics (age, gender, race, and ethnicity), alcohol abuse, non-opioid abuse, tobacco use, mental disorders, and hepatitis C were reported for the study population. For this study, mental health disorders were identified as depressive disorder (ICD-9 codes 296.2x, 296.3x, 300.4, 311), suicide attempt or other self-injury (ICD-9 codes E95x.x, E98x.x), or anxiety disorder (ICD9 codes 300.0x, 300.21, 300.22, 300.23, 300.3, 308.3, 309.81).

d. **Report Type Selection**

Due to the large number of report types generated by IUH systems, we focused on the most prevalent and relevant note types available in the database. To increase efficiency, the text mining process was limited to report types that were objectively deemed relevant to OUP identification through consulting clinical and data subject experts at Indiana University and the Regenstrief Institute. Three types of reports were initially excluded due to presumed irrelevance to OUP identification: radiology reports, medication fills, and patient instructions. Next, the study investigators selected 9 report types with assumed relevancy to OUP identification and queried the study population for their 2014 counts. The system returned 142,971 reports as the following: Emergency Department Doctor Progress Notes (48,898), Emergency Department Discharge Notes (28,637), Primary Care Doctor Outpatient Progress Notes (26,669), Visit Notes (21,868), Discharge Summary (11,731), History and Physical (2,759), Admission History & Physical (1,390), Preadmission History/Physical (576), Primary Care Doctor Outpatient History and Physical /Initial Consult (443). By exploring the 9 relevant report types, we found that the top 5 report types generated over 96% (137,802) of total reports, while the bottom 4 were only responsible for 4% of reports generated. Due to labor constraints, the bottom 4 reports (Admission History and Physical, Preadmission History/Physical, Primary Care Doctor Outpatient History and Physical /Initial Consult) were excluded. Thus, Emergency Department Doctor Progress Notes, Emergency Department Discharge Notes, Primary Care Doctor Outpatient Progress Notes, Visit Notes, Discharge Summary, were used for the text-mining analysis.

e. **Identify Patients with Opioid Use Problems**

We used two criteria which are explained in the following section to identify opioid use problem. The criteria will cover text-mining and ICD-9 approaches which are explained in details in the following sections.

Identify OUP using a Text Miming Process

The process of identifying OUP using text mining involved 2 main steps: 1) algorithm development to flag potential positive reports; and 2) results of the validation process using semi-assisted manual review.

1- Algorithm Development to Flag Potential Positive Cases Using nDepth™

We applied the text-mining package provided by nDepth™ to parse medical notes to detect OUP. nDepth™ is an NLP tool designed by the Regenstrief Institute in Indianapolis to extract data from the Indiana Network for Patient Care (INPC), which is a healthcare database managed by Regenstrief Institute on behalf of the Indiana Health Information Exchange (IHIE) [94,95]. As IUH is part of INPC, nDepth™ was used to query patients' clinical notes to identify possible opioid use problems. To develop the algorithm, 2 keyword lists adopted from the literature were entered into nDepth™ [91]. The first list was comprised of opioid terms (e.g., Vicodin, Opiate), and the second list was comprised of problem terms (e.g., addiction, abuse) (Appendix A.1). nDepth™ creates state machines, an algorithm that parses report types for certain criteria, to check for all possible combinations of the 2 lists within a 5-word distance. Flagged results are automatically checked for whether the statement is negated, hypothetical, historical, or experienced. This process itself was adopted from the ConText algorithm developed by Harkema et al. (2009) [96]. Thus, if the system determined the patient statement was not negated or deemed hypothetical or historical, those flags were deemed as experienced (considered positive).

2- Semi-Assisted Manual Review Validation Process

Flagged reports were semi-assisted manual reviewed by 2 trained reviewers. In case of disagreement, an expert physician acted as a third reviewer to resolve the dispute. To avoid overlooking any signs of opioid use problems, nDepth™ was programmed to highlight 27 suggestive phrases in the flagged reports. These phrases were collected from the literature and modified based on common clinical dialog in Indiana (Appendix A.2). In cases where there was more than one flagged report per patient, the system randomly selected one type of the flagged reports to be reviewed per patient. The criteria chosen to determine OUP were adopted from Carrell et al.s' (2015) work and listed in Table 3.1. Of note, as using marijuana medically and recreationally is illegal in Indiana, its use was treated as concurrent use of an illicit drug during the process of manual review. Other modifications were also adopted based on initial reviews of subsets of patients' clinical reports.

f. **Identify OUP Using ICD-9 Codes**

Two definitions from the literature utilizing ICD-9 codes were combined to create a case definition of OUP: opioid abuse and dependence (304.00, 304.01, 305.50, 305.51, 304.71, 304.02 , 304.70, 305.52) and opioid poisoning (965.0, 965.00, 965.01, 965.02, 965.09, E850.0, E850.1, E850.2) (Appendix A.3). These definitions were combined to minimize the chance of systematically creating type II errors by capturing the wider spectrum of opioid use problems among the study population.

Table 3.1 The criteria for identifying opioid use problems in patients' clinical notes

No.	Criteria for opioid use problems
1	Substance abuse treatment, including referral or recommendation
2	Methadone or suboxone treatment for addiction
3	Obtained opioids from nonmedical sources
4	Loss of control of opioids, craving
5	Family member reported patient's opioid addiction to clinician
6	Significant treatment contract violation
7	Concurrent alcohol abuse/dependence (not remitted)
8	Concurrent use of illicit drugs
9	Current or recent opioid overdose
10	Pattern of early refills (not an isolated event)
11	Manipulation of physician to obtain opioids
12	Obtained opioids from multiple physicians surreptitiously
13	Opioid taper/wean due to problems, lack of efficacy (not due to expected pain improvement)
14	Unsuccessful taper attempt
15	Rebound headache related to chronic opioid use
16	Concurrent use of unauthorized narcotics (polypharmacy)
17	Physician states opioid abuse/overuse/addiction or listed ICD codes for opioid abuse/dependence

g. **Analysis**

Frequency tables will be used to describe studied population characteristics. For inferential statistics, characteristics of positive OUP identified by text mining and ICD cohorts will be compared. Chi-square will be used for categorical data analysis and independent t-test will be used to compare means for continuous variables.

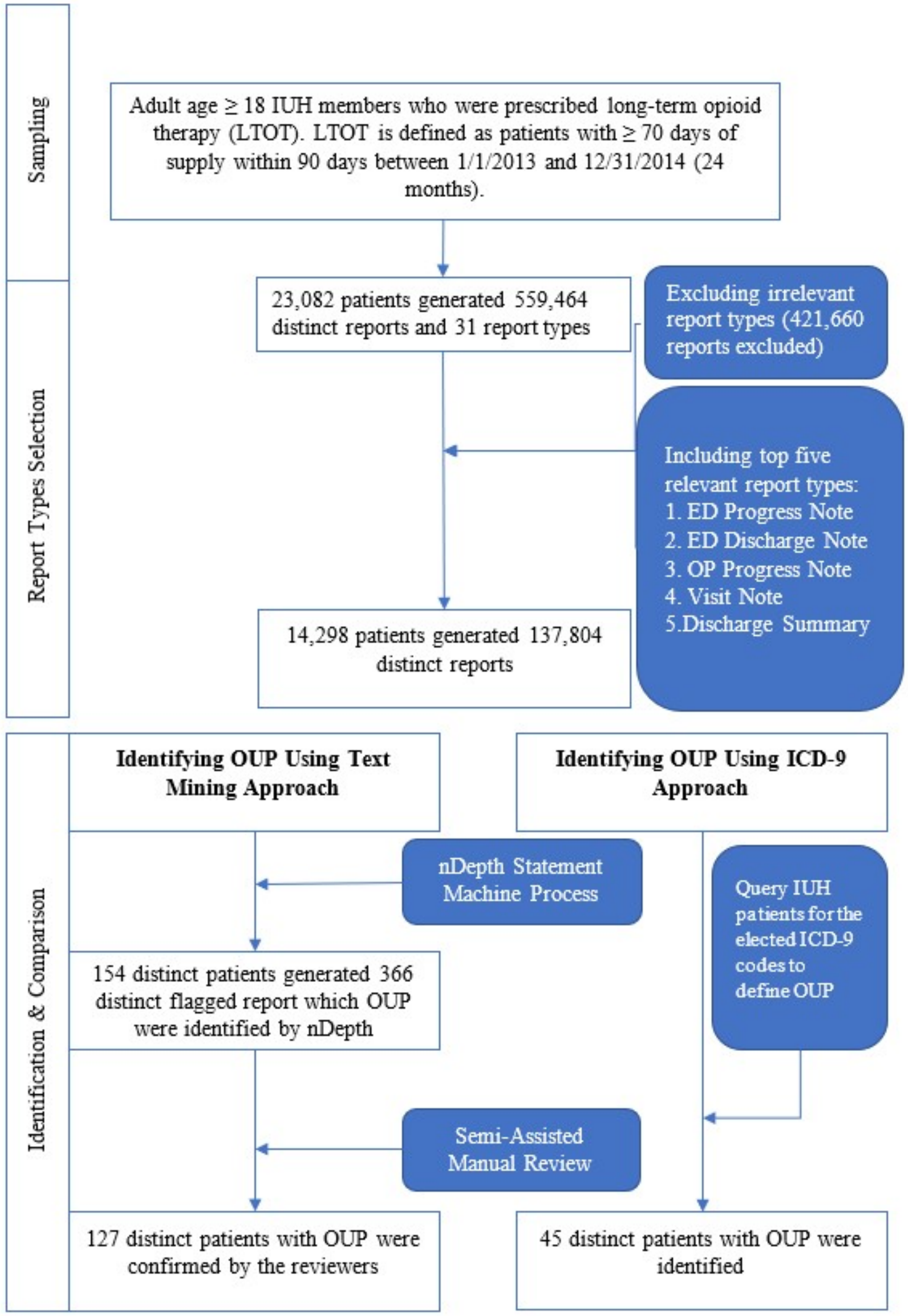


Figure 3.1 Identification of opioid use problems summary methods and results

C. Results

We identified and compared OUP using ICD-9 codes and text mining techniques. The results are summarized in Figure 3.1.

a. Sampling Results

Adult patients from IUH on LTOT were queried, and our inclusion criteria returned 34,661 patients. Our exclusion criteria, cancer and schizophrenia, excluded 11,579 (33.4%) patients (11,121 cancer patients, 272 with schizophrenia [186 with both schizophrenia and cancer]). We identified 23,082 patients who received LTOT and generated 559,464 reports across the IUH system. Our report selection criteria excluded 8,784 patients, leaving 14,298 eligible patients which generated 137,804 reports. Nearly 60% of the study population were over 55 years old, and women represented 62% of the study population. The white race was dominant among the study population (86%), while black patients represented 12%, and other or unknown races were 2%. Non-Hispanic or Latino ethnicity represented 87% of the study population (Table 3.2).

b. Identification of Opioid Use Problems Using a Text-Mining Approach

Our algorithm flagged 468 distinct statements in 366 unique reports representing 154 patients. For validation and to confirm the presence of OUP, 2 reviewers each reviewed 1 flagged report per positive patient based on specific criteria (Table 3.1). Out of 154 flagged patients, only 127 patients were deemed as positive cases of OUP with a positive predictive value of 82%. Cohen's κ was run to determine if there was agreement between 2 reviewers' judgement on whether a subset of 200 patients in the study cohort were meeting any OUP criteria. There was moderate agreement between the 2 reviewers judgements: κ = .0.691 (95% CI, 0.58 to 0.79), $p < .0005$.

Table 3.2 Demographics characteristics of study population

Demographics	**Study Cohort N(%)**
Age (18-24)	236 (1.7%)
25-34	891 (6.2%)
35-44	1,749 (12.2%)
45-54	2,808 (19.6%)
55-64	3,222 (22.5%)
>65	5,392 (37.7%)
Sex (women)	8,923 (62.4%)
Men	5,375 (37.6%)
Race (Black)	1,719 (12.0%)
White	12,367 (86.5%)
Other/Unknown	212 (2.0%)
Ethnicity (Not Hispanic or Latino)	12,369 (86.5%)
Hispanic or Latino	138 (1.0%)
Unknown	1,791 (12.5%)
Alcohol Abuse (Yes)	88 (0.6%)
Non-opioid Abuse (Yes)	88 (0.6%)
Tobacco Use (Yes)	912 (6.4%)
Depression (Yes)	845 (5.9%)
Self-injury (Yes)	14 (0.1%)
Hepatitis C (Yes)	94 (0.7%)
Total	**14,298**

Frequency distribution of criteria to identify OUP in patients' clinical notes were ranked as follows [OUP criteria, Frequency (percentage)]: [Physician states narcotic abuse/overuse/addiction or listed ICD codes for opioid abuse/dependence, 48 (38%)]; [Concurrent use of illicit drugs, 18 (14%); Methadone or suboxone treatment for addiction, 15 (12%)]; [Concurrent use of unauthorized narcotics (polypharmacy), 11 (9%)]; [Current or recent opioid overdose, 10 (8%)]; [Concurrent alcohol abuse/dependence (not remitted), 9 (7%)]; [Obtained opioids from multiple physicians surreptitiously, 8 (6%)]; [Physician or patient wants immediate taper, 3 (2%)]; [Substance abuse treatment, including referral or recommendation, 2 (2%)], [Significant treatment contract violation, 2 (2%)]; [Family member reported patient's opioid addiction to clinician, 1 (1%)]

c. **Comparison Between Text Mining and ICD-9**

In comparison to the text-mining approach, querying the same study cohort for designated ICD-9 codes to identify OUP returned 49 distinct events representing 45 unique patients (4 patients had 2 distinct positive OUP ICD-9 codes). The frequency distribution of positive ICD-9 codes for OUP were ranked as the following [ICD-9 code, Frequency (percentage)]: [304, 24 (49%), 305.5, 9 (18%)], [304.01, 5 (10%), 965, 3 (6%), 965.01, 3 (6%)], [E850.2, 2 (4%)], [304.71, 1 (2%)], [965.09, 1 (2%)], [E850.0, 1 (2%)] (Appendix A.4 and Appendix A.5).

The overlap between the 2 OUP positive cohorts was measured, and we found 8 cases that were positive in both a text mining approach and ICD-9 codes query (Appendix A.1). The frequency distribution of OUP positive cohorts characteristics are summarized in Table 3.3.

The frequency distribution percentage of positive OUP among stratified positive cohorts (ICD and text mining) by gender shows that women had a higher frequency among ICD cohorts compared to the text mining cohort (71% and 48%). However, the rate of overall positive OUP indicates a higher rate of positive OUP of men vs. women (14.7 vs. 10.4 per 1,000).

The frequency distribution percentage of positive OUP among stratified positive cohorts (ICD and text mining) by age group shows that the 45-54 age group had the highest frequency percentage of OUP (31% and 24%). However, the rate of positive OUP per group age indicates the 18-24 group age had the highest OUP rate among age groups in ICD and text mining cohorts (12.7 and 42.4 per 1,000) (Table 3.3 and Figure 3.2).

Table 3.3 Frequency distribution of OUP positive cohorts' characteristics

Characteristic	Study Cohorts* ICD-9 N (%)	Text-Mining N (%)	Combined N (%)	Significance X2 (P-value)**
Age (18-24)	3 (6.7%)	10 (7.9%)	13 (7.9%)	10.2 (< 0.001) †
25-34	7 (15.6%)	22 (17.3%)	26 (15.9%)	
35-44	4 (8.9%)	29 (22.8%)	33 (20.1%)	
45-54	14 (31.1%)	31 (24.4%)	42 (25.6%)	
55-64	7 (15.6%)	25 (19.7%)	31 (18.9%)	
>65	10 (22.2%)	10 (7.9%)	19 (11.6%)	
Sex (Women)	32 (71.1%)	61 (48.0%)	89 (54.3%)	7.1 (0.0076) †
Men	13 (28.9%)	66 (52.0%)	75 (45.7%)	
Race (Black)	5 (11.1%)	23 (18.1%)	27 (16.5%)	1.6 (0.6014)
White	39 (87.7%)	101 (79.5%)	134 (81.7%)	
Other/Unknown	1 (2.0%)	3 (2.4%)	3 (1.8%)	
Ethnicity (Not Hispanic or Latino)	39 (86.7%)	106 (83.5%)	139 (84.8%)	0.8 (1)
Hispanic or Latino	0 (0.0%)	2 (1.6%)	2 (1.2%)	
Unknown	6 (13.3%)	19 (15.0%)	23 (14.0%)	
Other Characteristics				
Alcohol Abuse (Yes)	2 (4.4%)	3 (2.4%)	5 (3.0%)	0.5 (0.6069)
Non-opioid Abuse (Yes)	10 (22.2%)	7 (5.5%)	14 (8.5%)	10.4 (0.0028) †
Tobacco Use (Yes)	14 (31.1%)	17 (13.4%)	30 (18.3%)	7 (0.0079)
Depression (Yes)	12 (26.7%)	14 (11.0%)	23 (14.0%)	6.3 (0.0118) †
Self-injury (Yes)	2 (4.4%)	2 (1.6%)	3 (1.8%)	1.2 (0.2804)
Hepatitis C (Yes)	3 (6.7%)	3 (2.4%)	5 (3.0%)	1.8 (0.1848)
Total Cohorts	**45**	**127**	**164***	**—**

*ICD-9: Positive OUP cases using ICD-9 codes; Text Mining: Positive OUP cases using a text mining approach; Combined: Combined positive cases in ICD-9 codes or a text mining approach (There are 8 cases that overlapped between ICD and text mining cohorts); (%): Column percentage. ** Chi-square was reported for 22 measurements and Fishers exact test (Freeman-Halton test) for r x c tables.
† Significant difference between ICD and Text-mining cohort on $\alpha = 0.05$.

Table 3.4 Comparison of care utilization among OUP positive cohorts

Care Utilization	ICD Positive Mean ± (SD)	Text-mining Positive Mean ± (SD)	P-value
Outpatient Visits	12.84 ±(15.05)	12.87 ± (12.402)	0.992
Emergency Department visits	1.91 ± (2.193)	3.17 ± (4.915)	0.022 †
Hospitalizations	2.24 ± (2.838)	1.55 ± (2.572)	0.154
Cumulative Hospitalizations Days	16.24 ±(23.985)	7.43± (12.016)	0.022 †

† Results are significant on $\alpha = 0.05$

Similar to the overall study population, race and ethnicity were dominantly White, non-Hispanic in both positive cohorts (Table 3.3). Other characteristics associated with OUP in the literature were reported, including alcohol abuse, non-opioid abuse, tobacco use, depression, self-injury, and hepatitis C. All showed a pattern of being at their lower point in the negative cohort, with a modest increase in the positive text-mining cohort and peaking at the positive ICD-9 cohort (Table 3.3). An independent-samples t-test was conducted to compare care utilization (outpatient visits, emergency department visit, hospitalizations, cumulative hospitalization days) among OUP positive cohorts (Table 3.4).Our analysis shows that positive text mining cohort had significantly higher average of visiting emergency department (M = 3.17, SD= 4.9) comparing to ICD positive cohort (M = 1.9, SD= 2.1), P = 0.022. The ICD positive cohort had a significantly higher average of cumulative hospitalization days (M = 16.2, SD= 23.9) comparing to the text-mining positive cohort (M = 7.4, SD= 12, P = 0.022.)

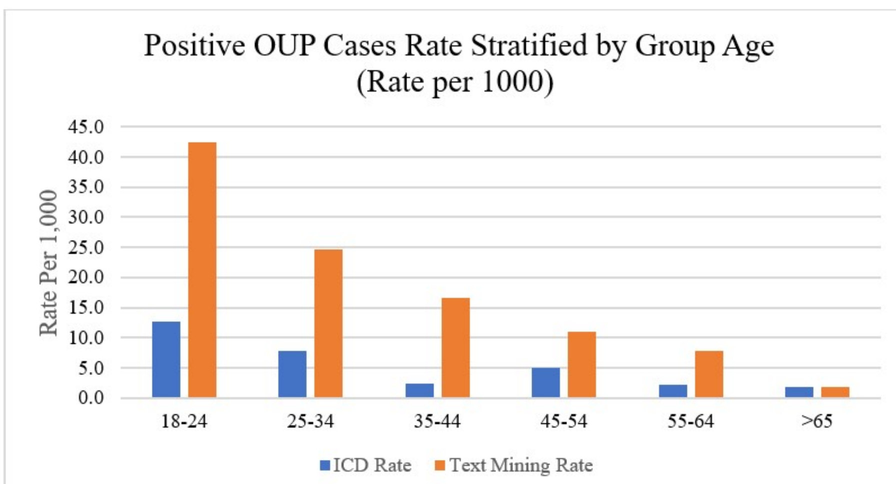


Figure 3.2 Positive opioid use problems rate stratified by age group

d. **Variables of Interest**

In this study, we conducted a bi-variate analysis to compare positive text-mining cohorts vs. negative cohort. Chi-square and likelihood ratio were reported for categorical variables (Table 3.5) and an independent t-test were reported for continuous variables (Table 3.6). The following variables were significant on alpha= 0.05 and were ranked based on likelihood ratio and approximate change in probability: Non-opioid Abuse, Gender, Tobacco Use, Self-Injury, Depression, Alcohol, Abuse, and Hepatitis C. The t-test was also significant on the following continuous variables: age, outpatient visits, emergency department visits, hospitalizations, cumulative hospitalization days.

Table 3.5 Measure of association and likelihood ratio for categorical variables

Variable	X2	p-value	Likelihood ratio	p-value	Approximate change in probability
Non-opioid Abuse	55.457	< 0.001	20.083	< 0.001	>45%
Gender	11.23	0.001	10.849	0.001	>45%
Tobacco Use	10.881	0.001	8.41	0.004	>40%
Self-Injury	30.971	< 0.001	7.947	0.005	>35%
Depression	6.186	0.013	4.994	0.025	>25%
Alcohol Abuse	6.616	0.01	3.823	0.051	>20%
Hepatits C	5.895	0.015	3.517	0.061	>20%

Table 3.6 Comparison of continuous variables among positive OUP using text mining vs. negative cohorts

Variable	Text-mining positive	Mean	Std. Deviation	p-value
Age	Yes	44.98	13.24	< 0.001
	No	59.02	16.34	
Outpatient Visits	Yes	12.86	12.4	0.001
	No	9.15	10.48	
Emergency Department visits	Yes	3.17	4.91	< 0.001
	No	0.94	1.86	
Hospitalizations	Yes	1.55	2.57	< 0.001
	No	0.54	1.138	
Cumulative Hospitalization Days	Yes	7.42	12.01	< 0.001
	No	3.21	17	

D. **Discussion**

We identified 23,082 patients who received LTOT during the study period, of which 14,298 have eligible text mining of relevant electronic clinical notes that yielded 127 positive OUP cases, compared to 45 cases using ICD-9 codes for the same population. Our study methodology followed the footsteps of previous work of Carrell et al. (2015), Palmar et al. (2015), and Hylan et al. (2015) [61,90,91]. This offers a unique opportunity to compare population characteristics and text mining results to identify opioid use problems in multi-healthcare settings. Below, we highlight takeaway findings of our study.

a. **Identification of Opioid Use Problems**

We measured the prevalence of opioid use problems in IUH adult patients using text mining and ICD-9 codes. In our study, validated-text mining identified 127 (0.8%) positive OUP cases out of 14,298 patients, while ICD-9 had identified 45 (0.3%) positive OUP cases from the same population (Total combined = 164 [1.1%]). Carrell et al. (2015)–validated NLP have identified 1,875 (8.5%) patients out of 21,795 patients eligible for the study, while ICD-9 had identified 2,240 (10.1%) from the same population. We believe the difference of opioid use problem prevalence between the 2 studies could be due to several factors such as the difference between opioid problem rates in both populations or the study design. In our study, we included 5 different reports types that covers emergency department and primary care visits, however, our study, as the opposite from Carrell's, did not have access to behavioral\mental health reports due to 42 CFR part 2 - confidentiality of substance use disorder patient records [97]. These reports could potentially include opioid use problem related data and adding them to our analysis could decrease the prevalence disparities between the 2 studies.

Despite the differences between the 2 combined percentages, both findings fall within the estimated range reported in the literature. According to Ballantyne's commentary published in the journal journal PAIN in 2015, published estimates of opioid use problems ranged from < 1% to 50% [98]. Our findings are consistent with other studies estimating opioid problems in the context of long-term opioid therapy. Noble et al. (2010) conducted a study to assess safety, efcacy, and effectiveness of opioids taken long-term for chronic non-cancer patients. The study reviewed 26 articles and included only patients who had at least 6 months of opioid treatment. According to the study, signs of opioid addiction were reported in only 0.27% of 4,893 study participants [99].

b. **Frequency of ICD Criteria to Identify OUP**

Our study used a broad-spectrum definition of opioid use problems comprised of 2 sets of ICD-9 codes for opioid abuse and dependence-as well as opioid poisoning. This is different from Carrell's study which only included those for opioid abuse (305.5, 305.51, or 305.52), and opioid dependence (304, 304.01, 304.02, 304.7, 304.71, or 304.72). Our inclusion of opioid poising codes has detected an additional 8 distinct patients. This equals 17% of the overall positive OUP detected by ICD-9 codes in our study.

Overall, our study population generated 49 ICD-9 events, representing 45 distinct patients. The most commonly used ICD-9 codes were 304 [opioid dependence unspecified] (49%), 305.5 [opioid abuse unspecified] (18%), and 304.1 [opioid dependence continuous] (10%). These findings are similar to the results reported by Palmer et al. (2015). The study investigated the prevalence of opioid problems among chronic opioid therapy patients in the Group Health Cooperative from 2006 to 2012. The most commonly reported ICD-9 codes were 304 [opioid dependence unspecified] (55%), 304.1 [opioid dependence continuous] (26%), and 305.5 [opioid abuse unspecified] (10%) [100]. Similar to the common use of unspecified ICD codes to identify OUP, a pattern was noticed by our reviewers during a manual review process of patients' clinical notes. In this study, we found unspecified criteria [Physician states opioid abuse/overuse/addiction or listed ICD codes for opioid abuse/dependence] represented 38% of positive OUP in the text mining approach. These findings suggest a common pattern in OUP identification between ICD-9 codes and a text-mining approach.

c. **Frequency of Text-Mining Criteria to Identify OUP**

In our study, we have used 17 criteria to identify OUP from clinical notes. The main difference in these criteria compared to Carrells' are the following: 1- we have included marijuana as part of our list of illicit drugs 2- We added concurrent use of authorized narcotics to include specific cases of polypharmacy of narcotic abuse. We added a criteria to cover cases where physician states narcotic abuse, overuse, addiction or listed ICD codes for opioid abuse/dependence. In our study, use of an illicit drug criteria comprised of 18 cases, of which 3 cases indicated active marijuana use (within the last month) without mentioning other illicit drug abuse. Polypharmacy of narcotic abuse flagged another 11 OUP cases. The third criteria yielded 48 positive OUP cases, of which, 16 cases have one or more mentions of ICD-9 code diagnosis of abuse and dependence within the clinical notes. Out of those 16 patients, there was only one patient who had a documented ICD-9 for opioid abuse and dependence in the structured data.

Demographics

In this study, we found women were represented more than men, overall. This is consistent with Carrell et al. (2015) and Hylan et al. (2015) which, both have found women to represent about two-thirds of the chronic opioid therapy population in their studies. Overall, higher women's representation is consistent with the national estimates of opioid prescriptions per gender. According to the national statistics by the CDC, there were 58% women vs. 42 % men who filled at least one opioid prescription in 2014[95]. However, despite more women than men seeking opioid prescriptions in the U.S., men had higher rates of opioid-related disorders, such as opioid abuse or opioid overdose. In 2016, the opioid overdose deaths by gender showed men had 18.1 deaths per 100,000 vs. 8.5 for women in the U.S.and 16.7 vs. 8.5, respectively, in Indiana [96]. Our study results are consistent with this notion; while positive OUP distribution stratified by gender showed a higher rate of women with OUP, the rate of overall positive OUP indicates a higher rate of positive OUP for men.

Mirroring nationwide trends of opioid prescription by age, age distribution of the overall study cohort was skewed toward an older age (> 65), which was the highest representation in the study cohort [97]. Frequency of positive OUP distribution was the highest in the group age 45-54 in both the ICD and text-mining methods. However, the rate of positive OUP per group age indicates younger age groups had the highest OUP rate among age groups in ICD and text-mining cohorts. This should inform physicians in Indiana that despite the higher frequency of older patients with opioid problems that they might encounter, younger-age patients are the ones at higher risk for an opioid problem.

A Note About Opioid Problem Detection Using Structured Data

Despite the study which used a combination of 2 types of ICD-9 codes (opioid abuse and dependence and poisoning), overall, our text-mining approach identified more cases. The text-mining approach also detected 15 patients with a mention of an ICD-9 code of opioid abuse and dependence in patients' clinical notes, but these codes were undocumented in patients' record as structured data. This might indicate a significant lack of documentation toward opioid use problems and confirm previous studies' findings regarding the under-reporting of OUP using ICD-9. For future research pertaining to identifying OUP, we recommend using alternate methods (in addition to ICD codes), such as text mining and machine learning, and further exploring other commonly documented structured data, such as procedural codes and electronic lab results.

d. **Limitations**

In this study, we used a text-mining approach to analyze patients' clinical notes to identify OUP. However, the analysis was limited to 5 report types. Developing and implementing NLP techniques generally requires intensive computing and an initial commitment of substantial resources and expertise [101]. Thus, it was not financially feasible within the allocated budget for this study to test all report types generated by the IUH study population. Rather, the authors relied on expert opinions to meaningfully limit the number of report types. Reliance on expert opinion during algorithm development is common in opioid problem identification literature. Canan et al. (2017) reviewed 15 automated algorithms to identify nonmedical opioid use using electronic health record data. The study found that investigators explicitly relied on subject matter experts during the process of algorithm development and to identify candidate variables [101].

Another limitation pertaining to study results is that, despite text mining correctly identifying 127 positive cases out of 154 initial positive cases (PPV = 82%), the overlap between the text-mining results and ICD-9 positive results was only 8 cases (17%) (Appendix A.1). This may suggest that the text mining method has missed some positive cases (false negative). To investigate this assumption, we have reviewed 500 randomly selected reports that were not flagged by our text-mining criteria. Among the 500 reports, we identified 10 false negative cases, indicating a negative predictive value of 98% on the reviewed subset.

Finally, we used multiple behavioral concepts to define opioid use problems (dependence, abuse, and addiction) and thus, distinguishing specific behavioral concepts was not automated in this research. Future work may investigate using text mining, natural language processing, and machine learning to better target specific behaviors.

E. **Conclusion**

In this study, we developed a text-mining approach to identify OUP among long-term opioid therapy patients at IUH. Compared to ICD-9 codes, text mining successfully identified more OUP cases within the study population. Future development of text-mining techniques can help identify cases undiscovered by conventional ICD-9 reporting methods.

4. MACHINE-LEARNING APPROACH TO IDENTIFY OPIOID USE PROBLEMS

A. Introduction

In 2012, opioid prescriptions reached their peak with 81.3 prescriptions per 100 people [98]. Despite more recent reports showing that this number has decreased to 66.5 prescriptions per 100 people, the opioid prescription rate in 2015 remains 3 times higher than the rates from 1999 [99], and more than a quarter of U.S. counties are still showing dispensed opioid prescription numbers equivalent to or higher than their populations [98]. The disparities over prescribing opioids across the nation has continued throughout the years. In fact, in 2016, some counties had prescription rates 7 times higher than the nation's average [98–100].

The link between average daily opioid prescription and opioid overdose has been established [101]. According to Dunn et al. (2010), long-term opioid therapy patients receiving higher doses of opioids are at a higher risk of drug overdose [102]. Other studies have found an association between heroin use and an increased rate of nonmedical opioid use and meet the criteria for opioid abuse [103].

Despite the importance of the opioid abuse issue, the current healthcare system mainly relies on the International Classification of Disease (ICD) to report use problems, which underestimates the actual number of patients exhibiting the target categorization [57,58].

In recent years, several studies have been conducted to investigate the validity of other methods, such as natural language processing and text mining, to measure opioid problems, in addition to ICD codes. However, to our knowledge, no studies have been conducted to measure the presence of opioid problems among long-term opioid users via machine-learning techniques. The aim of this chapter is to investigate the application of machine-learning techniques to classify opioid use problems (OUP) utilizing patients clinical notes. For this study, we define OUP as the presence of opioid aberrant behaviors, opioid misuse, opioid abuse and opioid use disorder in the patient' clinical notes.

B. Method

a. Machine-Learning Process Summary

To prepare for machine-learning process, 2 reviewers had reviewed 700 clinical notes using nDepth™ to determine opioid use problem cases. Report contents, along with review results, were extracted by the Regenstrief Institute data core into comma-separated values (CSV) files and Portable document format (PDF) files. The review of 700 clinical notes resulted in 145 positive reports for opioid use problems (OUP) and 555 negative reports with no indication of OUP in the text. To ensure HIPAA compliance, the files were stored and processed on the Karst desktop, an IU remote secure supercomputer (Figure 4.2).

To prepare data for machine learning, 2 Java programs were written to extract text from each of the PDF and CSV clinical notes files and they were saved into separate TXT files. For the machine-learning process, Waikato Environment for Knowledge Analysis (WEKA) was used to build and compare machine-learning models on a document level. Performance was compared based on F-measure, recall, and precision for 5 models. The best model performance was optimized and finalized into a standalone product to identify the OUP outcome among long-term opioid therapy patients at IUH based on their medical records. Finally, the model was tested on a subset of unseen data, and the model reported the results (Figure 4.2).

Gold Standard Development (nDepth™)

To create a gold standard for the machine-learning process, 2 trained reviewers manually reviewed 700 reports with a semi-assisted manual review process using nDepth™. The Regenstrief Institute in Indianapolis designed nDepth™ as a natural language processing tool with which to extract data from the Indiana Network for Patient Care (INPC). Furthermore, nDepth™ was programmed to highlight 27 suggestive words in flagged reports (Table A.2). When more than one report per patient was flagged, the system randomly selected one type of the flagged reports per patient to be reviewed. To establish the Kappa coefficient, both reviewers reviewed 200 identical reports. In case of the discrepancy, files were reevaluated and consensus was reached. Cohen's κ was run to determine if there was agreement between the 2 reviewers judgment regarding whether a subset of 200 patients in the study cohort met any OUP criteria. There was moderate agreement between the 2 reviewers judgments: κ = 0.691 (95% CI, 0.58 to 0.79), p < .001. Our overall gold standard dataset consisted of 136 positive OUP reports and 480 negative reports. The criteria to determine opioid use problems in clinical notes are listed in Table 3.1.

Data Transformation (Java)

The 700 manually reviewed clinical notes, which consisted of 132 PDF and 568 CSV reports, were uploaded into Karst. Karst Desktop is a remote desktop service for users with accounts on the Karst research supercomputer [92]. As part of the machinelearning processes, text from both CSV and PDF reports needed to be transformed into TXT files for each report. Thus, a Java program was written to extract data from the PDF and CSV files (Appendix D.1 and Appendix D. 2). A sample of the extracted data was reviewed to validate the extraction process to confirm data correctness and completion. As a final step to prepare the data for WEKA, text was converted into Attribute-Relation File Format (ARFF). ARFF is "an ASCII text files that describes a list of instances sharing a set of attributes. ARFF files were developed by the machine-learning project at the department of computer science of the University of Waikato for use with the WEKA" [103].

C. **Analysis**

a. **WEKA Configuration and Model Comparison**

In the process of model building, training the model on a larger sample size usually produces higher model performance compared to training the model on a smaller sample size [104]. However, the model-building process also considers class representation balance in the training process, [105,106] which can lead to the inability to classify the unbalanced data due to lack of prior information in the model training [106,107]. Class imbalance is characterized as there being "many more instances of some classes than others" [107,108]. The fundamental challenge with training a model on imbalanced data is the ability of imbalanced data to "significantly compromise the performance of the most standard learning algorithms" [109].

In previous years, many solutions have been introduced to deal with the imbalanced data problem and can be summarized as follows: 1- Data sampling: in which the model is trained on a more balanced distributed dataset [107,110]; 2- Algorithmic modification: in which adopting base-learning algorithms needs "to be more attuned to class imbalanced" [107,111]; 3- Cost-sensitive learning: this approach considers modification on data level, algorithm level, or both, in which considering misclassification of one data class (e.g., positive class) cost is higher than another data class (e.g., negative class) thus, algorithm performance is adjusted to minimize the higher cost error (misclassify positive class) [107,112].

Our gold standard dataset has about an 80% negative and a 20% positive class imbalance. Balancing the data in the WEKA training model on 50/50 or 40/60 class representation is recommended [113]. However, adopting this approach might jeopardize the real-world application of the model, given the difference in prior probability between the training environment and real-world application [107,114]. Hence, for this chapter, we decided to use an ensemble approach, where first we explore multiple algorithms and various classification configurations and compare their performance on a balanced subset of data that has similar representation between positive and negative reports. Second, once we determine the best WEKA configuration with which to classify the 2 classes, we will optimize the model by introducing 560 (80%) of the gold standard reports to enhance the aspect of real-world application of the model (because real world data is not balanced) [111]. To standardized testing environment, cross-validation (k = 10) will be used across all of the models.

Finally, the optimized model will be tested on a unseen data subset consisting of 140 (20%) clinical notes. Our intention in the optimization step is to maximize positive report detection (less false negatives), while maintaining minimal loss of precision. The rationale behind the ensemble approach is that classifiers are commonly designed to minimize error when they make predictions about new data. However, this assumption is proper when the cost of different errors is equally important [112,115]. In this research, we sought to prioritize the minimization of false negatives (by capturing more true positives, while risking an increase in capturing false positives). Accuracy measures, such as F-measures, recalls, and precision, will be reported for each step. The highlights of the WEKA configuration (Table 4.1) and the complete model schema (Figure 4.1) for best model performance will be reported. The complete model training steps of WEKA are reported in Appendix C.1 to Appendix C.19.

```
“Scheme: weka.classifiers.meta.FilteredClassifier -F
&quot;weka.filters.unsupervised.attribute.StringToWordVector -R first-last -W 1000 -prune-
rate -1.0 -T -I -N 0 -L -stemmer weka.core.stemmers.IteratedLovinsStemmer -
stopwords-handler weka.core.stopwords.MultiStopwords -M 1 -tokenizer
\&quot;weka.core.tokenizers.NGramTokenizer -max 3 -min 1 -delimiters \\\&quot;
\\\\r\\\\n\\\\t.,;:\\\\\\\&#39;\\\\\\\&quot;()?!\\\&quot;\&quot;&quot; -S 1 -W
weka.classifiers.functions.SimpleLogistic --
-I 0 -M 500 -H 50 -W 0.0”
```

Figure 4.1 WEKA model configurations

Table 4.1 Highlight of WEKA configurations

Option	Chosen Features	Comment
Classifier	FilteredClassifer	Cost-sensitive classifier was added to optimize the final model.
Algorithm	1-Linear Algorithm: Simple Logistic 2-Non-Linear Algorithm: SMO /J48/NaiveBayes 3-Ensemble Algorithm: RandomForest	For Simple Logistic: batch size:100, heuristic stop: 50, max boosting iteration: 500, number of boosting iteration: 0 number of decimal places: 2
Filter	StringToWordVector:	IDFTransform: True, TFTransform: True, LowerCaseToken: True. Words to keep: 1000
StopWordsHandler	MultiStopWords	
Tokenizer	NGramTokenizer	Max 3, Min 1

b. **Model Finalization**

The model with best accuracy measure performance will be saved as standalone MODEL files. The MODEL extension is a common files type for digital product definitions and simulations [116]. The files can be loaded by WEKA onto any PC or Mackintosh environment and used to classify cases of opioid use problems in unseen reports from IUH.

D. Results

a. Model Comparison

We used WEKA to build and test files among different models to classify IUH clinical notes into positive and negative for opioid use problems using a balancedclass dataset. The training set consisted of 136 positive reports and similar negative reports, which were uploaded to WEKA. The following algorithms were used: J48 (default WEKA algorithm), simple logistic, naive bayes, random forest, and sequential minimal optimization (SMO). All models were tested based on a 10-fold crossvalidation. Our model comparison has shown the Simple Logistic algorithm showed better performance than other models for predicting positive class (Precision: 91.9%; Recall: 91.2%; F-measure: 91.5%; ROC: 92%) (Table 4.2). The detailed accuracy measure performances for simple logistic are listed in Table 4.3.

b. Model Optimization

To enhance model real-world application, simple logistic model configuration was optimized on 80% of the gold standard dataset, which was comprised of 560 reports (136 positive and 424 negative). Compared to the balance dataset, the model had initially underperformed when larger portions of the negative dataset were introduced (Table 4.4). The model has classified 37 (27%) out 136 positive cases as false negative cases.

Table 4.2 Comparison of multiple models' performances to identify positive OUP

Accuracy measures to classify positive reports			
Algorithm	Precision	Recall	F-Measure
Simple Logistic	0.919	0.912	0.915
J48	0.847	0.853	0.85
RandomForest	0.756	0.934	0.836
SMO	0.797	0.838	0.817
NaiveBayes	0.821	0.64	0.719

Table 4.3 Detailed accuracy performance measure for simple logistic algorithm to classify OUP

Class	TP Rate	FP Rate	Precision	Recall	F- Measure	ROC Area
Positive	0.921	0.079	0.919	0.912	0.915	0.931
Negative	0.921	0.088	0.915	0.921	0.918	
Weighted Avg.	0.917	0.083	0.917	0.917	0.917	

Table 4.4 Initial simple logistic model performance on 80% of the gold standard data

Class	TP Rate	FP Rate	Precision	Recall	F- Measure	ROC Area
Positive	0.728	0.052	0.818	0.728	0.77	0.94
Negative	0.948	0.272	0.916	0.948	0.932	
Weighted Avg.	0.895	0.219	0.892	0.895	0.892	

Table 4.5 Optimized simple logistic model performance on 80% of the gold standard data

Class	TP Rate	FP Rate	Precision	Recall	F- Measure	ROC Area
Positive	0.897	0.094	0.753	0.897	0.819	0.890
Negative	0.906	0.10 3	0.965	0.906	0.934	
Weighted Avg.	0.904	0.101	0.913	0.904	0.906	

Table 4.6 Simple logistic confusion matrix comparison before and after optimization

Confusion Matrix (N=560: 136 Positive, 424 Negative)			
Before optimization		After Optimization	
Negative	Positive	Negative	Positive
402	22	384	40
37	99	14	122

Table 4.7 Simple logistic model performance tested on 20% of unseen gold standard data

Class	TP Rate	FP Rate	Precision	Recall	F- Measure	ROC Area
Positive	0.889	0.030	0.667	0.889	0.762	0.924
Negative	0.970	0.111	0.992	0.970	0.981	
Weighted Avg.	0.965	0.106	0.971	0.965	0.967	

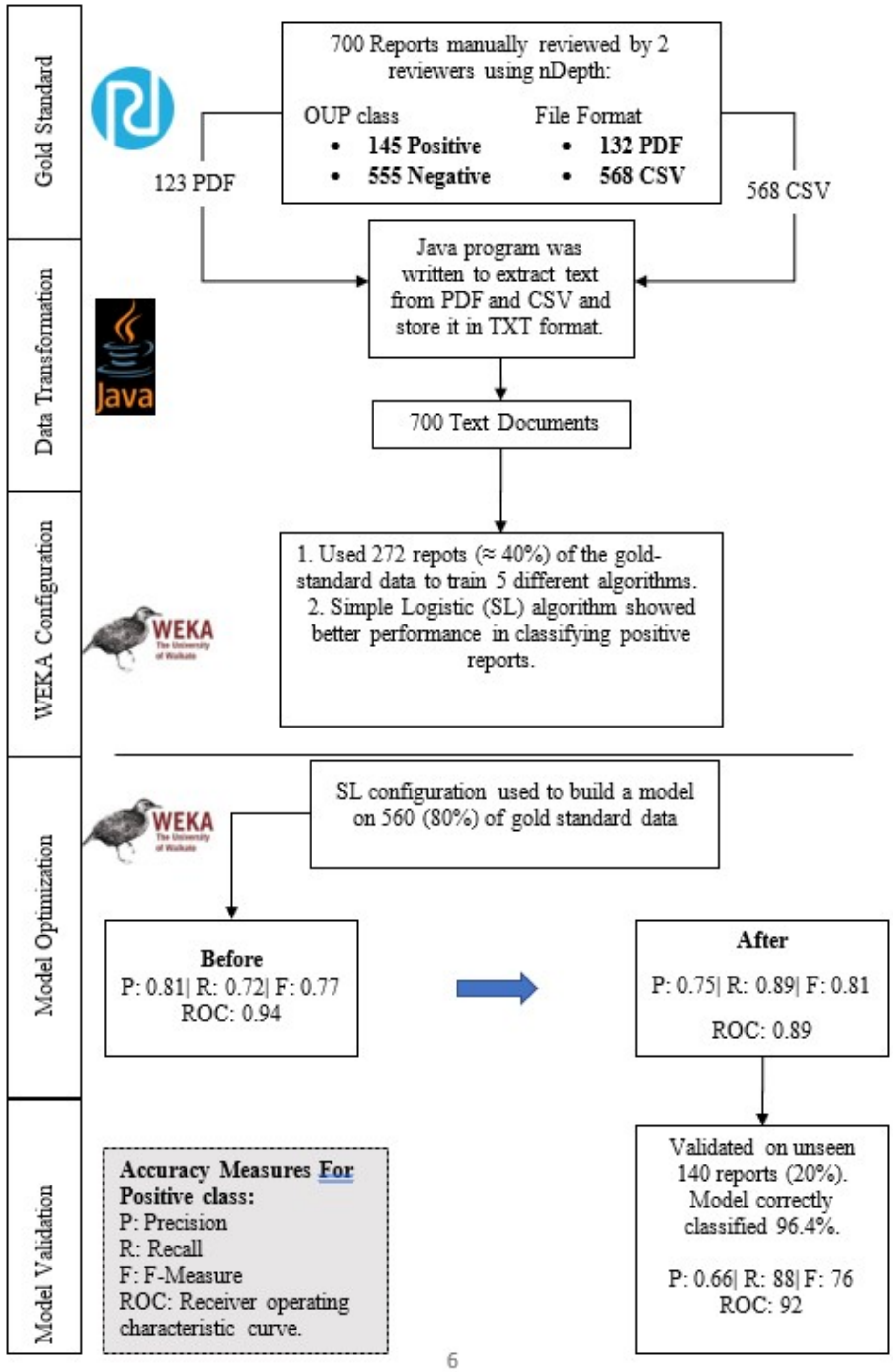


Figure 4.2 Summary of chapter 4 method and results

To account for this effect, a cost-sensitive classifying technique was adopted. Costsensitive classifying is one method to fine-tune model performance through increasing the penalty of misclassifying one or more classes in the confusion matrix

(Appendix C.20 to Appendix C.28). Adjusting the model cost sensitively matrix to 2:1 has doubled the penalty of misclassifying positive cases as false positives, which resulted in a decrease in false positive cases (Table 4.5), dropping the total to only 14 (10%) cases out of 136. A complete confusion matrix for both models' outcomes is listed in Table 4.6.

c. **Model Testing**

The finalized model was tested on 140 reports (20% of our gold standard data). These data were not used initially in the process of training and testing the model. On the unseen data, the model has correctly classified 136 (96.4%) reports with 88.9% sensitivity to identify positive cases. The accuracy measures to classify positive. reports were the following: Precision: 0.667, Recall: 0.889, and F-Measure: 0.762. The complete performance measures are listed in Table 4.7

E. **Discussion**

In this chapter, we built a machine-learning model to identify opioid use problems based on Indiana University Health clinical notes. The positive and negative reports, which had been extracted and manually reviewed using nDepth™ in the previous chapter, were pre-processed and uploaded into WEKA, which was used to create a model to identify opioid use problems. In this section, we highlight some of the key points of this work.

a. **Algorithms' Performance**

We tested 5 different algorithms to classify opioid use problems on a document level (Appendix B.1 to Appendix B.5). simple logistic regression showed overall higher accuracy performance measures in classifying positive opioid use problems on a balanced dataset (Table 4.2). We found this interesting, given the growing number of published articles discussing the superiority of random forest and/or support vector machines over logistic regressions [117,118]. However, a recent article by Pranckeviˇcius et al. (2017) compared Naive Bayes, Random Forest, Decision Tree, Support Vector Machine and logistic regression classifiers to classify text reviews (multi-class textual data). The study found that logistic regression has achieved the highest classification accuracy compared with other classifiers [119]. This can be explained by the signal-to-noise ratio. One opinion article we found discussed the advantages of logistic regressions over tree decision models; the article found that logistic regressions perform better when the signal-tonoise ratio is low [120].

Given the large text data (noise) compared to the signal (phrases or statements indicating an opioid use problems), we decided to investigate this hypothesis and see if logistic regression will still perform better if we limit our model training to sentences and phrases. Thus, we have manually reviewed all of the positive reports and 300 of the negative reports to annotate statements and phrases that indicate or negate the presence of opioid use problems using the Extensible Human Oracle Suite of Tools (eHOST). This process has resulted in 255 sentences and phrases indicating the presence of an opioid use problems. However, it only yielded 55 sentences and phrases, negating the presence of an opioid use problems. Finally, we applied similar settings to compare the following 4 classifiers: logistic regression, random forest, Naive Bayes and SMO. Our results showed that random forest and SMO scored slightly higher than logistic regression and Naive Bayes (Appendix B.6).

b. **Cost-Sensitive Classifier**

In this study, we found that cost-sensitive classifier is an effective method to use to adjust for the data imbalance. Adjusting the cost matrix has increased recall performance while lowering the precision of predicting positive cases in the model by increasing (doubling) the penalty for making a false negative error. Our rationale for this approach arises from the intention to minimize missing true-positive cases. We believe that identifying more positive OUP cases will allow future researchers to better understand why this sector of positive OUP patients (according to their reports) was not documented using ICD as an opioid use problems case. Furthermore, due to the importance of opioid issues from a clinical perspective, and given the small representation of opioid use problems cases in the IUH system (about 1%), building a high-sensitivity model to detect positive OUP in long-term opioid therapy patients based on their clinical notes may seem like a practical approach, given the small percentage of positive OUP cases. Such a pragmatic approach might not be uncommon in screening tests; mammograms are a common example of screening tests where sensitivity is higher than specificity [121,122].

c. **Limitation**

Our gold standard data were determined by manually reviewing the text-mining algorithm to identify opioid use problems. Any positive report subclass that was not captured by the text-mining algorithm may not be represented in our model. To account for this limitation, we have intentionally added all false positive reports as part of our training model and labeled them as negative reports. Furthermore, 500 random negative reports (from the text-mining algorithm) were manually reviewed, and false negative reports were added to the training set as positive cases.

F. **Conclusion**

In this chapter, we adopted an ensemble approach to build a machine-learning model to identify opioid use problems based on Indiana University Health clinical notes. Our final model was tested on unseen data and reported a sensitivity of 88% when identifying positive cases. We concluded that a machine-learning approach may be used to identify opioid use problems in Indiana.

5. SUMMARY

A. Summary Overview

In this thesis, we aimed to build a classification model to identify opioid use problems using text-mining and machine-learning approaches. Additionally, we examined several opioid use problems classification tools/models and identified their strengths and weaknesses. Generally, these tools can be categorized into three classes, depending on the data population method as follows: 1- Self-report tools (First generation), 2- Structured data models (Second generation), 3- Clinical notes models (Third generation). Below we synthesize the findings of this thesis in light of this classification.

B. Self-report Tools (First Generation)

These are risk-assessment tools that predict the risk of current or future opioid use problems based on predictive variables extracted from data self-reported by the patient or physician. Common examples of self-report tools are SOAPP, ORT, and COMM. The main advantage of these tools is their independence from electronic health record data and can be administered to the patient or filled out by the treating physician. However, having the majority of these tools solely rely on patient completion might have limited their ability to predict aberrant drug-taking behavior due to reporting bias or a lack of patient comprehension. Other factors that might hinder using self-report tools are the need for documentation from the medical staff and the large variation in its sensitivity across different populations and different outcome definitions.

C. **Structured Data Models (Second Generation)**

These are risk-assessment models that predict the risk of current or future opioid use problems based on predictive variables extracted from structured data of electronic health records. Common examples of these models are the work of Edlund et al. (2007), White et al. (2009), Rice et al. (2012), Turner et al. (2014), and others. These models do not rely on patient reported data (directly), rather, it depends on the structured data available in the electronic health records. However, this dependence on electronic health records structured data makes these models as good as the quality and robust nature of captured data stored in the electronic record. For example, the number of prescribers and the number of dispensers have been shown to be a risk indicator for opioid use problems. However, including these data into a classifying or predictive model may require pre-work to consolidate these data from multiple databases (such as Prescription Monitoring Programs). Thus, it was no surprise to see Age and Gender as the only common variables across the 9 models we found using structured data. However, with the constant growth of the Health Information Exchange in size and number, we expect to see more robust models in the future that identify and utilize new variables to better understand opioid use problems.

D. **Clinical Notes Models (Third Generation)**

a. **NLP and Text Mining**

These are risk assessment models that identify the presence of opioid use problems through classifying patients' clinical notes using natural language processing and text-mining approaches. In comparison to the 2 previous methods, NLP and text mining are more complex techniques that require intensive computational and labor skills. Thus, in our review, only a handful of studies were found using these techniques to identify opioid use problems. In this thesis, we used a text-mining approach to identify opioid use problems using patient's clinical notes. Below we highlight the findings of our research within the context of similar research in the literature.

Methods and Variables

Hylan et al. (2015), Carrell et al. (2015), and the subsequent work of Palmer et al. (2015) are a few models found that use NLP to identify opioid use problems from patients' clinical notes. Hylan et al. (2015), a two-year prospective study, used 3 phases of NLP to identify clinical notes with indications of 1- opioid overuse, misuse, abuse, or addiction 2- opioid dependence, and 3- "mentions" of ICD-9 codes (12 ICD-9 codes) for opioid abuse or dependence in the clinical notes. The study has identified 158 (5.7%) patients with opioid use problems, while ICD-9 for the same cohort has identified 25 (0.9%).

Our study used dictionary-based NLP which was adopted from Carrell et al. (2015) for 2 reasons. First, as the opposite of Carrell study's, Hylan et al. (2015) did not specify the technical details of their NLP process. Second, the NLP method of Carrell study's have resulted in detecting a larger percentage of opioid use problems. Third, Hylan et al. (2015) conducted a prospective study and it was limited to primary care settings, while Carrell et al. (2015) conducted a retrospective study and included primary care and emergency department notes. To note, both studies drew samples from the Group Health population. The major difference was in the study period (2 years vs. 5 years), study design (prospective vs. retrospective), and study sample (primary care vs. primary care and emergency). In our study, the validated-text mining identified 127 (0.8%) positive OUP cases out of 14,298 patients while ICD-9 has identified 45 (0.3%) positive OUP cases from the same population (total combined = 164 (1.1%)).

In addition to the NLP classification, both studies have included ancillary analysis of study cohort to estimate risk. Hylan studied prediction ability of 7 variables to identify a positive NLP cohort. These variables were documented for the 2 years prior to initiation of chronic opioid therapy. These variables were: Age (15 to 44 or 45 to 64), smoking status, and 2-year prior ICD codes for the following diagnoses: opioid abuse/dependence, drug abuse/dependence (excluding opioids), alcohol abuse/dependence, mental health disorder, and hepatitis C. The study found that using multivariate logistic regression, variables indicating age 18 to 44, opioid abuse diagnosis, positive smoking status, and mental disorder diagnoses were significant predictors during the measured period. This study was limited to primary care patients of an integrated group practice, salaried physicians working in Group Health Cooperative (GHC) clinics.

In our study, validated-text mining identified 127 (0.8%) positive OUP cases out of 14,298 patients, while ICD-9 identified 45 (0.3%) positive OUP cases from the same population (total combined = 164 [1.1%]). We compared demographic characteristics and variables of interest among 2 positive groups (text mining and ICD) using Chisquare and independent t-test. We have also compared positive OUP using text mining vs. negative OUP for the same variables and reported multivariate odds ratio. Our variables of interest included non-opioid abuse, gender, tobacco Use, self-Injury, depression, alcohol abuse, hepatitis C, age, outpatient visits, emergency department visits, hospitalizations, and cumulative hospitalization days. These variables were chosen subjectively based on our systematic review of previous work and the feasibility of obtaining these variables from IUH and INPC databases. Multivariate analysis logistic regression indicates that only gender, non-opioid abuse, self injury, outpatient visits, emergency department visits and inpatient visits were significant predictors to positive text-mining outcome (Table 5.1).

Table 5.1 Multivariate logistic regression of text-mining measure of opioid use problems

Variables	Odds Ratio	Lower*	Upper*
Depression	1.166	0.612	2.225
Age Continuous	0.915	0.868	0.965
Age Group	1.465	0.854	2.512
Sex	1.62	1.124	2.335
Race	0.928	0.591	1.456
Ethnicity	1.202	0.931	1.553
Alcohol Abuse	1.869	0.536	6.519
Other Medical Abuse	3.583	1.409	9.116
Tobacco Use	0.919	0.473	1.784
Self Injury	7.884	1.35	46.028
Hepatitis C	1.38	0.348	5.468
Outpatient visits	1.016	1.003	1.03
ED visits	1.093	1.053	1.135
Inpatient Visits	1.34	1.138	1.579
Hospital Days	0.983	0.958	1.009

*Odds ratio value based on 95% confidence interval

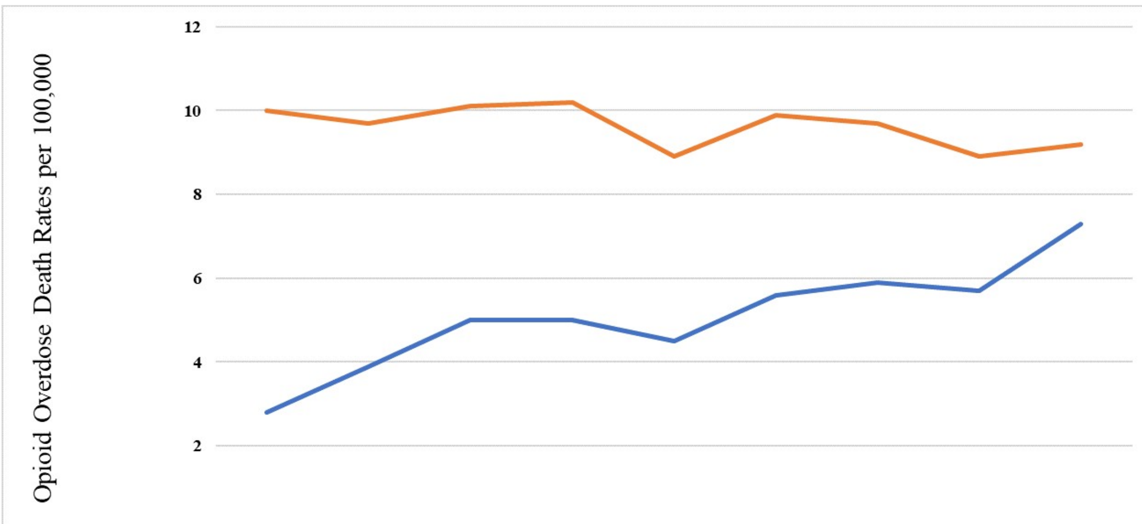
Collectively, in addition to younger age, prior opioid diagnosis and mental disorder, our study findings indicate non-opioid abuse, self injury, outpatient visits, ED visits and inpatient visits are significant variables to predict opioid use problems in positive NLP and text-mining positive cohort. Our reported bi-variate measure of association has also found alcohol abuse, hepatitis C, tobacco use, depression are significant associated with the measured outcomes. One common limitation of both studies was the small number of positive outcome (127 out of 14,298 and 158 out of 2,752), which affected the model significance to predict the outcome. This could be Due the limitations in the process of text mining, itself, and inability of dictionary-based method to detect indication of opioid use problems in indirect or long statements. Such statements might be deemed false positive.

Prevalence of Opioid Use Problems

We measured the prevelence of opioid use problems in IUH adult patients using text mining and ICD-9 codes. In our study, validated-text mining identified 127 (0.8%) positive OUP cases out of 14,298 patients, while ICD-9 has Identified 45 (0.3%) positive OUP cases from the same population (total combined = 164 [1.1%]). Carrell et al (2015) Validated-NLP method has identified 1,875 (8.5%) patients out of 21,795 patients eligible for the study, while ICD-9 has identified 2,240 (10.1%) from the same population. the disparities between the 2 results was vast in both the ICD and text mining/NLP approaches despite both studies using the same eligibility criteria of long-term opioid therapy (or chronic opioid therapy). This difference could be attributed to one or more of the following factors:

First, despite both studies using the same inclusion criteria, our study excluded patients with schizophrenia due to documented higher prevalence of OUP among this population. However, this factor would have minimal effect, since only 86 schizophrenic patients (who were already excluded due to the presence of cancer diagnosis) were additionally excluded from the study.

Second, this difference could be attributed to the time and study period of the 2 studies. While the longest available follow-up period in our study was 2 years, Carrell's study assessed patients over a 5-year period from 2006-2012. This could indicate that the length of time for opioid consumption would lead to an increase in prevalence of opioid problem. One critique to this assumption is that neither study measured the direct effect of the length of opioid therapy on development of the study outcomes.



	2006	2007	2008	2009	2010	2011	2012	2013	2014
Indiana	2.8	3.9	5	5	4.5	5.6	5.9	5.7	7.3
Washington	10	9.7	10.1	10.2	8.9	9.9	9.7	8.9	9.2

Figure 5.1 Opioid overdose death rates per 100,000 population from 2006-2014. Source: Kaiser Family Foundation [123]

Third, this could be attributed to the difference between opioid problem rates between the states of Washington and Indiana during the studied period for both studies. National statistics of opioid overdose death rates (per 100,000) shows that the rate of opioid overdose deaths in Washington state ranged from 10.2 to 9.2 during Carrell's study period (2006-2012), while the same rate ranged from 5.7 to 7.3 in Indiana during our study period (2013-2014) (Figure 5.1).

Finally, in the opposititION of Carrell's study, our study did not have access to behavioral\mental health reports due to 42 CFR Part 2 - confidentiality of substance use disorder patient records regulations [97]. We believe that including behavioral health reports is likely to identify more positive cases of opioid use problems in our study population.

Collectively, both studies used NLP to identify opioid use problems in primary care and emergency settings and compared the results with ICD-9 codes. The NLPtext mining technique identified additional positive cases in both studies. Text-mining approach also detected 15 patients with a mention of ICD-9 code of opioid abuse and dependence in patients' clinical notes but these codes were undocumented in patients' record as structured data. This might indicate a significant lack of documentation for opioid use problems and confirm previous studies' findings regarding the underreporting of opioid problems using ICD codes. For future research pertaining to identifying OUP, we recommend using alternate methods (in addition to ICD codes), such as text mining and machine learning, and further exploring other commonly documented structured data, such as procedural codes and electronic lab results.

b. **Machine Learning**

These risk assessment techniques utilize a set of statistical techniques to identify parts of speech, entities, sentiment, and other aspects of text. The chosen technique can be expressed in models that can be applied to another text [124]. These models can be used to identify a particular outcome in the clinical notes. These models are more flexible; to improve performance by training the model on larger and more representative training sets and by choosing a suitable classifying algorithm. In general, the learning process of machine learning can be categorized into supervised and unsupervised learning. Supervised learning means training documents are tagged with a specific class (e.g., positive, negative), while unsupervised, can extract meaning without a specific training set. A common example of unsupervised training is data clustering.

Machine Learning in Classification of Disease

In our review, we did not find any published work pertaining to machine-learning models that classify clinical notes to identify opioid-addiction related behavior. However, we found several studies which have used machine-learning algorithms to classify other medical events such as chronic disease hospitalization, or particular medical condition such as heart failure stage, cancer diagnosis or acute pancreatitis [125–128].

Despite the differences in the outcome of interest, the general theme of building a machine-learning model is similar. First, identify the outcome of interest and source of clinical notes (e.g. heart failure diagnosis in the emergency department discharge note). Second, identify training corpus (gold standard) using NLP/text-mining approach (e.g. rule-based). Third, build machine-learning model and evaluate the performance (by reporting F-measure, recall and precision).

Machine-Learning to Identify Opioid Use Problems

In this thesis, similar steps were adopted to build machine-learning predictive models to identify opioid use problems among long-term therapy patients at IUH using clinical notes. We built on our previous effort in text mining to investigate the application of machine-learning techniques to classify opioid use problems utilizing patients clinical notes. We manually reviewed a list of positive and negative clinical notes that resulted from the text-mining study, which we then used to build our machine-learning model. We tested 5 different algorithms to classify opioid use problems at the document level. Simple Logistic regression showed overall higher accuracy performance measures in classifying positive opioid use problems on balanced dataset. Our final model was tested on unseen data and reported a recall of 88% and precision of 66% (F-measure = 76%) when identifying positive cases.

Despite our findings the results of identifying opioid use problems are encouraging, our outcome of interest (OUP), comprised of multiple behavioral concepts, such as dependence, abuse, addiction, and overdose. While distinguishing specific behavioral concepts was not automated in this research, future work may investigate the use of text mining, natural language processing, and machine learning to better target specific behaviors before considering using these for surveillance purposes or incorporating them into a clinical decision system for primary care and emergency encounters.

E. **Limitations**

In this thesis, we used a text-mining approach to analyze patients clinical notes to identify OUP. Due to labor constraints, our analysis was limited to 5 report types, despite our effort to rely on expert opinions to meaningfully limit the number of report types. This approach might have systematically overlooked a certain group of positive OUP patients. Since our gold standard data was made through manually reviewing a text-mining algorithm to identify opioid use problems, any positive-report subclass that was not captured by the text-mining algorithm may not be represented in our model. To account for this limitation, we have intentionally added all falsepositive reports as part of our training model labeled as negative reports. Further, a random sample of 500 negative reports (by the text-mining algorithm) were manually reviewed and false-negative reports were added to into the training set as positive cases.

a. **Conclusion**

This study's main contribution was to demonstrate the validity of using a text-mining approach and machine-learning techniques to identify opioid use problems from patients' clinical notes in multi-site medical care facilities. Our research has also highlighted the main demographic characteristics, care utilization, and co-morbidity similarities and differences among opioid use problems identified by text mining versus the opioid use problems identified by ICD codes. Additionally, this research has compared multiple machine-learning algorithms' performance to identify opioid use problems from patients' clinical notes and has reported their performance to inform future research.

F. **Future Direction and Recommendations**

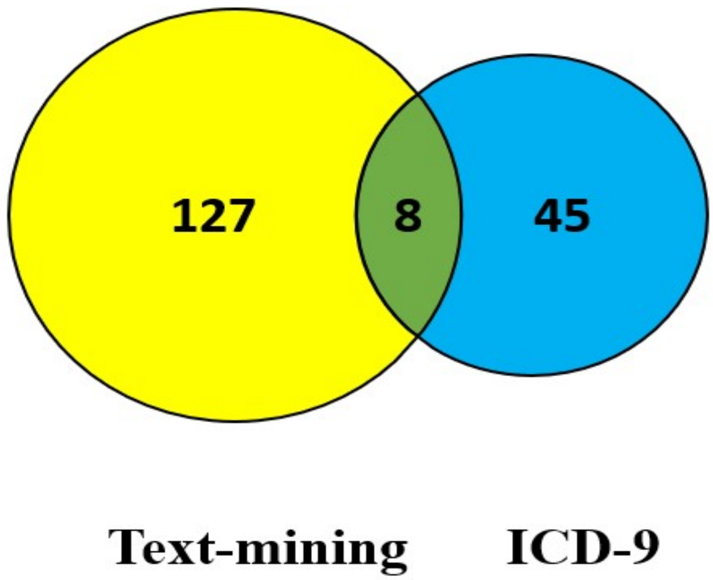
Based on the results from this thesis, future research may focus on:

1. Further analyzing machine-learning capability to identify opioid use problems from clinical notes. We recommend to build machine-learning model that is based on larger validated training corpus which comprised of multiple years of patient' clinical notes.

2. Comparing text-mining approach against ICD-10 to understand first, whether ICD-10 documentation to OUP has changed from ICD-9 and second, to further study the pattern, if any, for the OUP cases missed by ICD codes.
3. Exploring the idea of adopting a hybrid approach that combines ICD codes in addition to patients clinical notes (text mining or machine learning). Incorporating such approach may synergize the ability of future models to identify the opioid issue.
4. Focusing on measuring the relationship between opioid use problems and the length of opioid treatment in long-term opioid therapy patients.
5. Investigating the use of text mining, natural language processing and machine learning to better target specific opioid use problems behaviors (such as overdose vs. abuse) before incorporating these techniques into surveillance purposes or adopting it into a clinical decision system for primary care and emergency encounters.

6. APPENDICES

A. CHAPTER 3 SUPPLEMENTS



Appendix A.1 Venn diagram of the overlap between the two positive cohorts of opioid Use problem in the study

Appendix A.2 Terms used by nDepth™ algorithm to identify (flag) OUP positive reports

Opioid terms		**Problem use terms**		
codeine	analgesic	addict	abusing	aberrancy
codeine	narcotic	addicted	misusing	aberrant
codeine	opiate	addiction	overusing	daiversion
fentanyl	opioid		over using	divert
hydrocodone			abuse	
methadone			abuser	
morphine			misuse	
oxy			misuser	
oxy-ir			overuse	
oxycod			overuser	
oxycodone			over use	
oxyContin percocet polypharmacy vico Vicodin			over user	

Appendix A.3 Highlighted phrases used in the semi-assisted manually review process

Semi assisted manual review highlighted words		
abus*	opioid use disorder	diversion/divert
misus*	dependent	high
overus*	dependency	euphoria
over us*	heroin	abnormal urine drug screen
Addict*	Crav*	dirty urine
inconsistent urine	Lost prescriptions	alcohol abus*
unauthorized dose increase	Less suggestive words	early refills
taking more than prescribed	marijuana	immediate taper
Narcan	narcotic	naloxone

Appendix A.4 ICD-9 codes used to define OUP

Class Codes	
Diagnosis Category	**ICD-9-CM Codes**
Opioid Abuse and dependence	1. 304.00 (opioid dependence unspecified) 2. 304.01 (opioid dependence continuous) 3. 305.50 (opioid abuse unspecified) 4. 305.51 (opioid abuse continuous) 5. 304.71 (opioid/other dependence continuous) 6. 304.02 (opioid dependence episodic) 7. 304.70 (opioid/other dependence unspecified) 8. 305.52 (opioid abuse episodic) 9. 304.72 (opioid/other dependence episodic)
Poisoning	10. Non-specific code 965 Poisoning by analgesics antipyretics and antirheumatics 11. Non-specific code 965.0 Poisoning by opiates and related narcotics 12. Specific code 965.00 Poisoning by opium (alkaloids), unspecified 13. Specific code 965.01 Poisoning by heroin 14. Specific code 965.02 Poisoning by methadone 15. Specific code 965.09 Poisoning by other opiates and related narcotics 16. E850.0 Accidental poisoning by heroin 17. E850.1 Accidental poisoning by methadone 18. E850.2 Accidental poisoning by other opiates and related narcotics

Appendix A.5 Frequency distribution of positive ICD-9 codes for OUP

Frequency distribution of positive ICD-9 codes for OUP		
ICD-9 Codes	**Description**	**N (%)**
304	opioid dependence unspecified	24 (49%)
305.5	opioid abuse unspecified	9 (18%)
304.01	opioid dependence continuous	5 (10%)
965	Non-specific code 965.0 Poisoning by opiates and related narcotics	3 (6%)
965.01	Specific code 965.01 Poisoning by heroin	3 (6%)
E850.2	E850.2 Accidental poisoning by other opiates and related narcotics	2 (4%)
E850.0	E850.0 Accidental poisoning by heroin	1 (2%)
965.09	Specific code 965.09 Poisoning by other opiates and related narcotics	1 (2%)
304.71	opioid/other dependence continuous	1 (2%)
Grand Total		**49 (100) ***
*4 out of 45 patients had two positive ICD-9 codes for OUP		

B. **CHAPTER 4 SUPPLEMENTS**

Appendix B.1 J48 model performance on 80% of the gold standard data

Class	TP Rate	FP Rate	s **Precision**	Recall	F-Measure	ROC Area
Positive	0.691	0.092	0.707	0.691	0.699	0.808
Negative	0.908	0.309	0.902	0.908	0.905	
Weighted Avg.	0.855	0.256	0.854	0.855	0.855	

Appendix B.2 Naive Bayes model performance on 80% of the gold standard data

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Positive	0.625	0.146	0.578	0.625	0.601	0.851
Negative	0.854	0.375	0.877	0.854	0.865	0.776
Weighted Avg.	0.798	0.319	0.804	0.798	0.801	0.794

Appendix B.3 Random forest model performance on 80% of the gold standard data

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Positive	0.485	0.160	0.493	0.485	0.489	0.662
Negative	0.840	0.515	0.836	0.840	0.838	
Weighted Avg.	0.754	0.429	0.752	0.754	0.753	

Appendix B.4 SMO model performance on 80% of the gold standard data

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Positive	0.721	0.083	0.737	0.721	0.729	0.819
Negative	0.917	0.279	0.911	0.917	0.914	
Weighted Avg.	0.870	0.232	0.869	0.870	0.869	

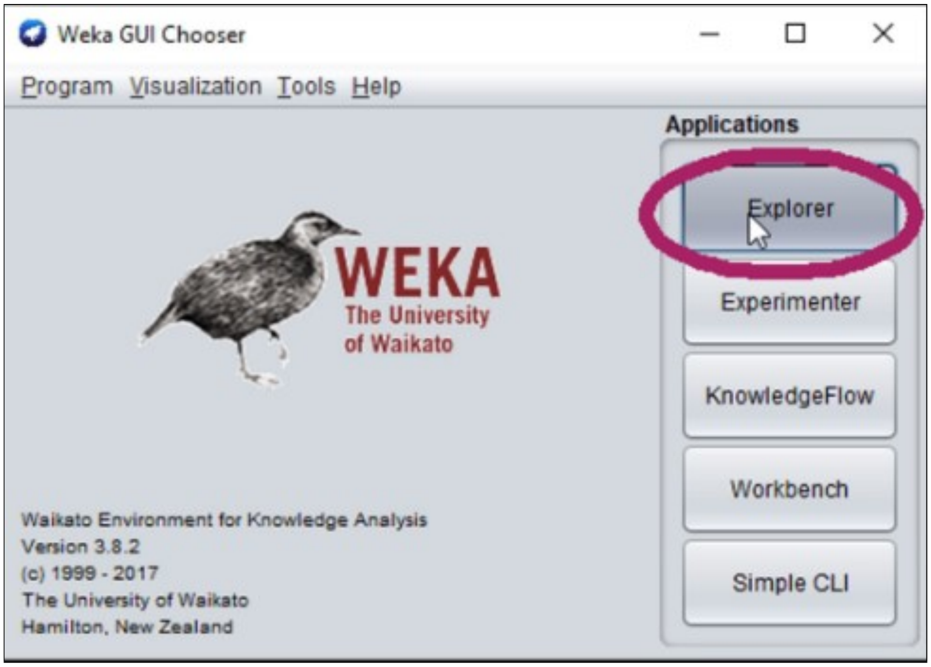
Appendix B.5 Simple logistic model performance on 80% of the gold standard data

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Positive	0.728	0.052	0.818	0.728	0.770	0.940
Negative	0.948	0.272	0.916	0.948	0.932	
Weighted Avg.	0.895	0.219	0.892	0.895	0.892	

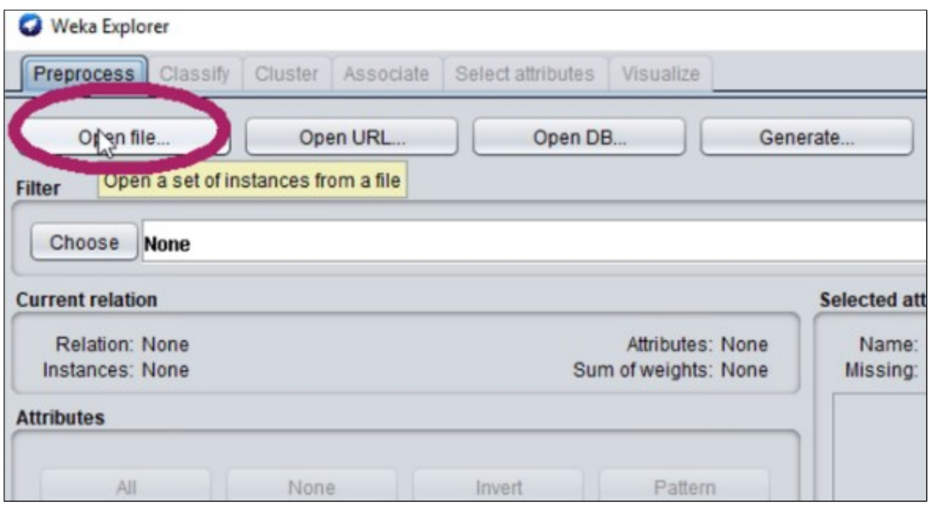
Appendix B.6 A list of document, sentence and phrase performance measure to classify positive class

		Accuracy Measures		
Level	**Algorithm**	**Precision**	**Recall**	**F-Measure**
Document	SimpleLogistic	0.918	0.904	0.911
	J48	0.847	0.853	0.85
	RandomForest	0.756	0.934	0.836
	BinarySMO	0.797	0.838	0.817
	NaiveBase	0.821	0.64	0.719
Sentence	RandomForest	0.867	1	0.929
	BinarySMO	0.867	1	0.929
	SimpleLogistic	0.856	1	0.922
	NaiveBase	0.856	1	0.922
Phrase	RandomForest	0.864	1	0.927
	BinarySMO	0.864	1	0.927
	SimpleLogistic	0.859	1	0.924
	NaiveBase	0.859	1	0.924

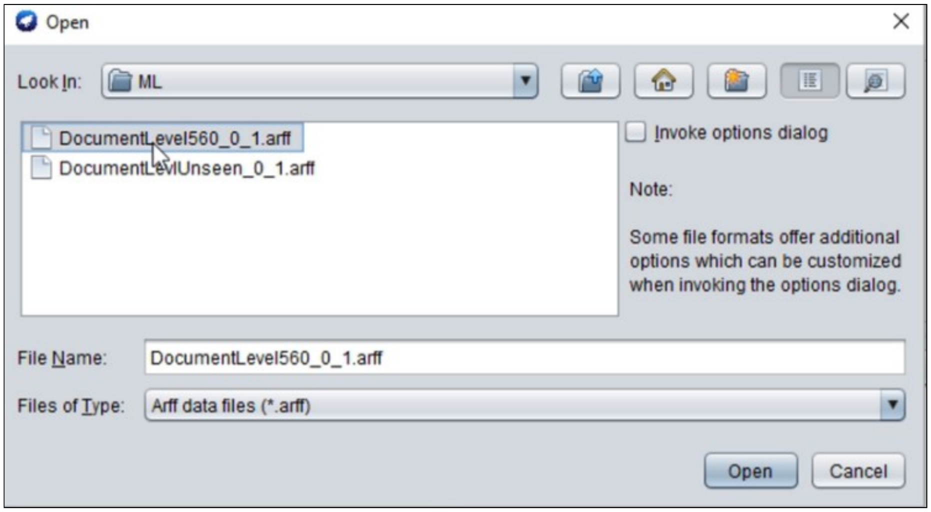
C. **WEKA DEMONSTRATION**



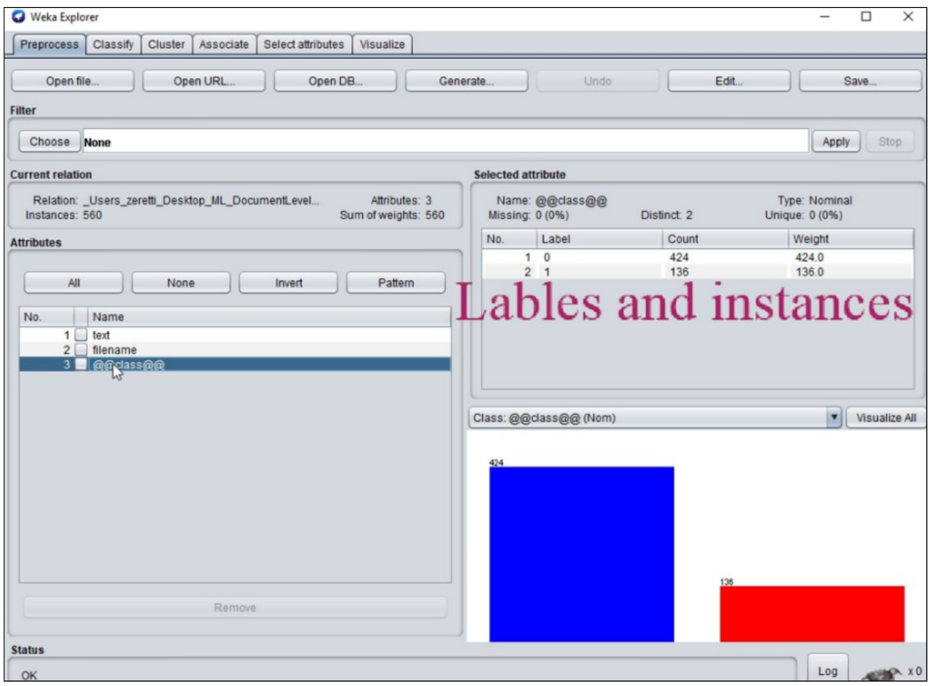
Appendix C.1 Click Explorer to open WEKA



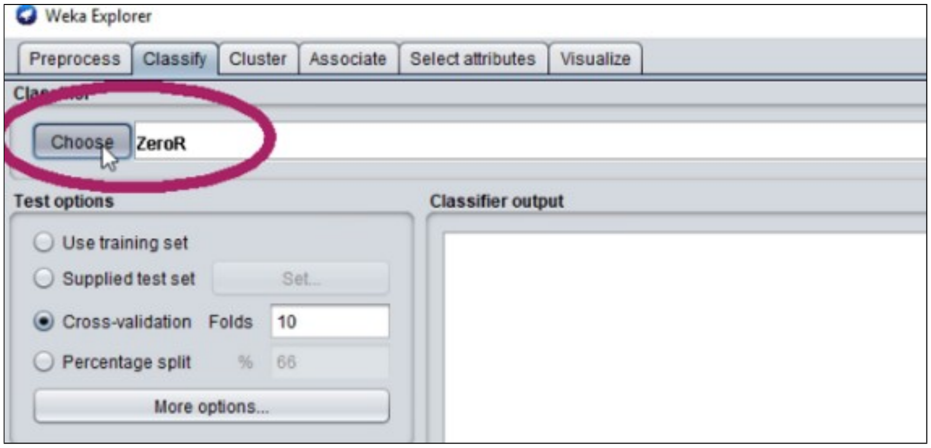
Appendix C.2 Click Open file



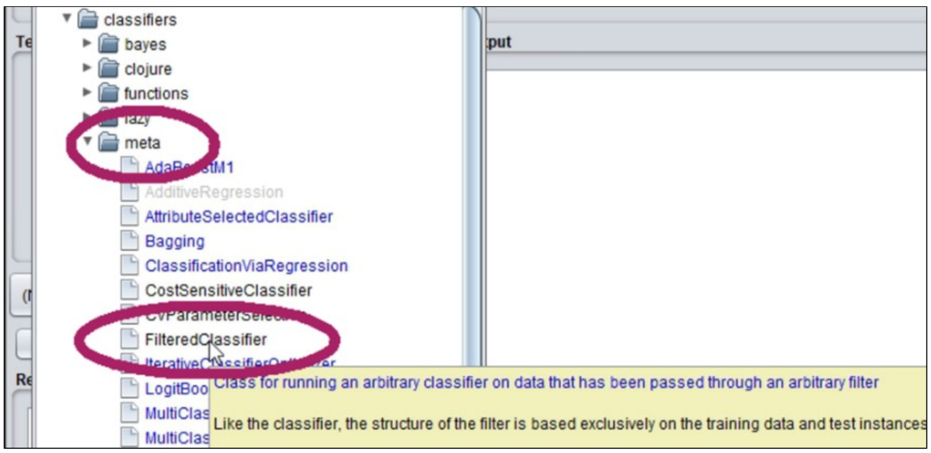
Appendix C.3 Choose ARFF training data file from computer



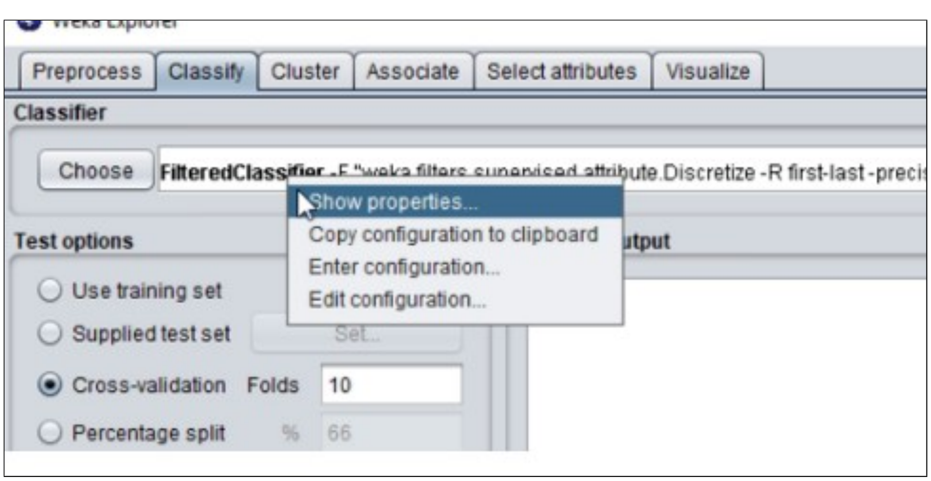
Appendix C.4 Double- check data before proceeding



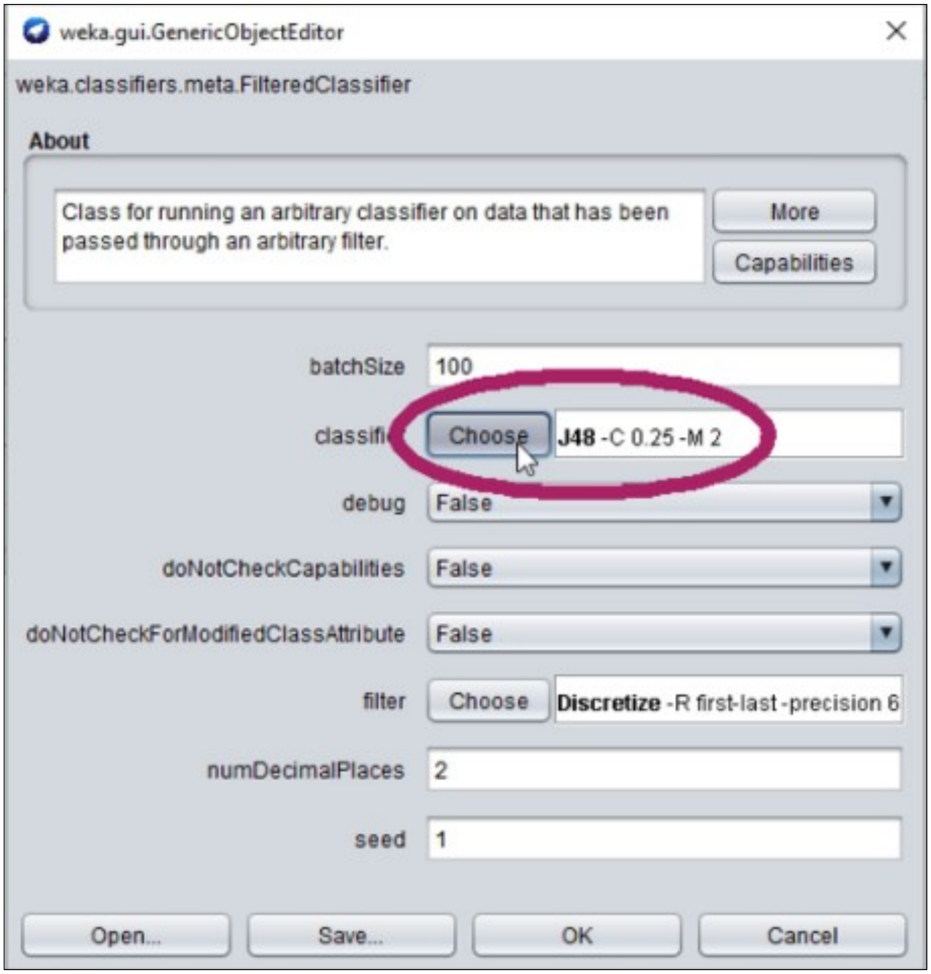
Appendix C.5 Click Choose to open classifier options



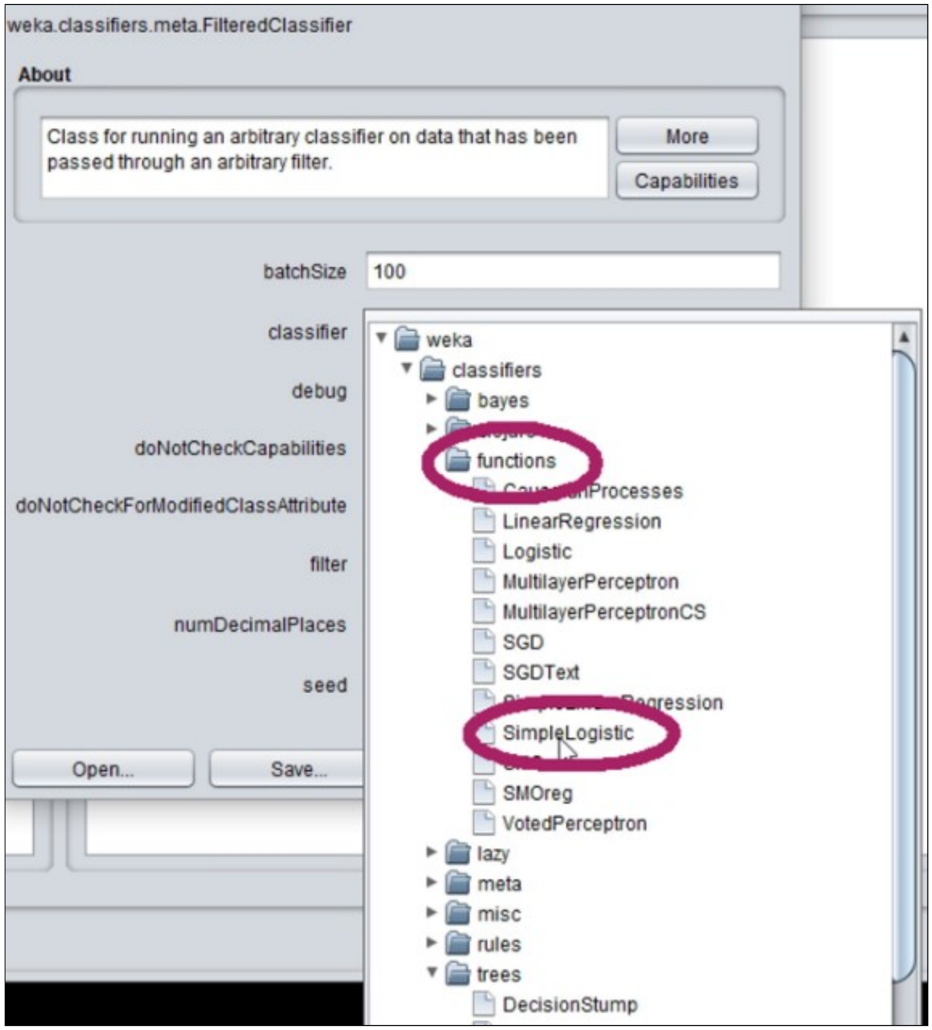
Appendix C.6 Choose filtered classifier



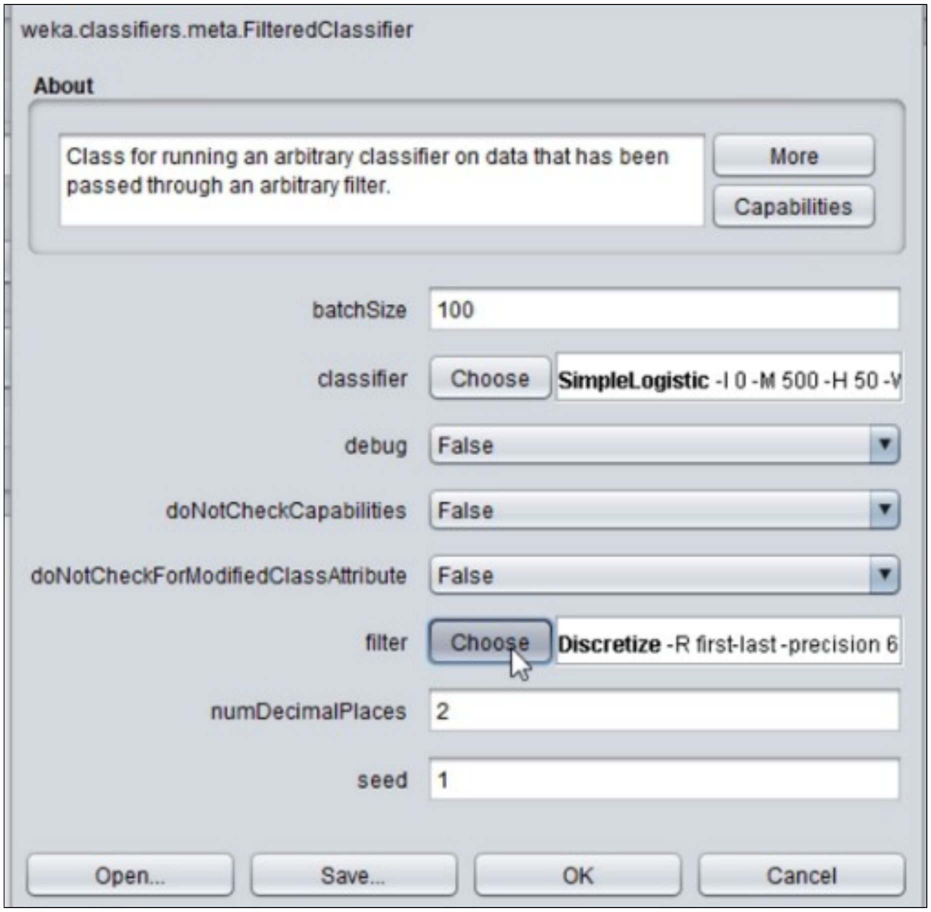
Appendix C.7 Right click on classifier name and click show properties



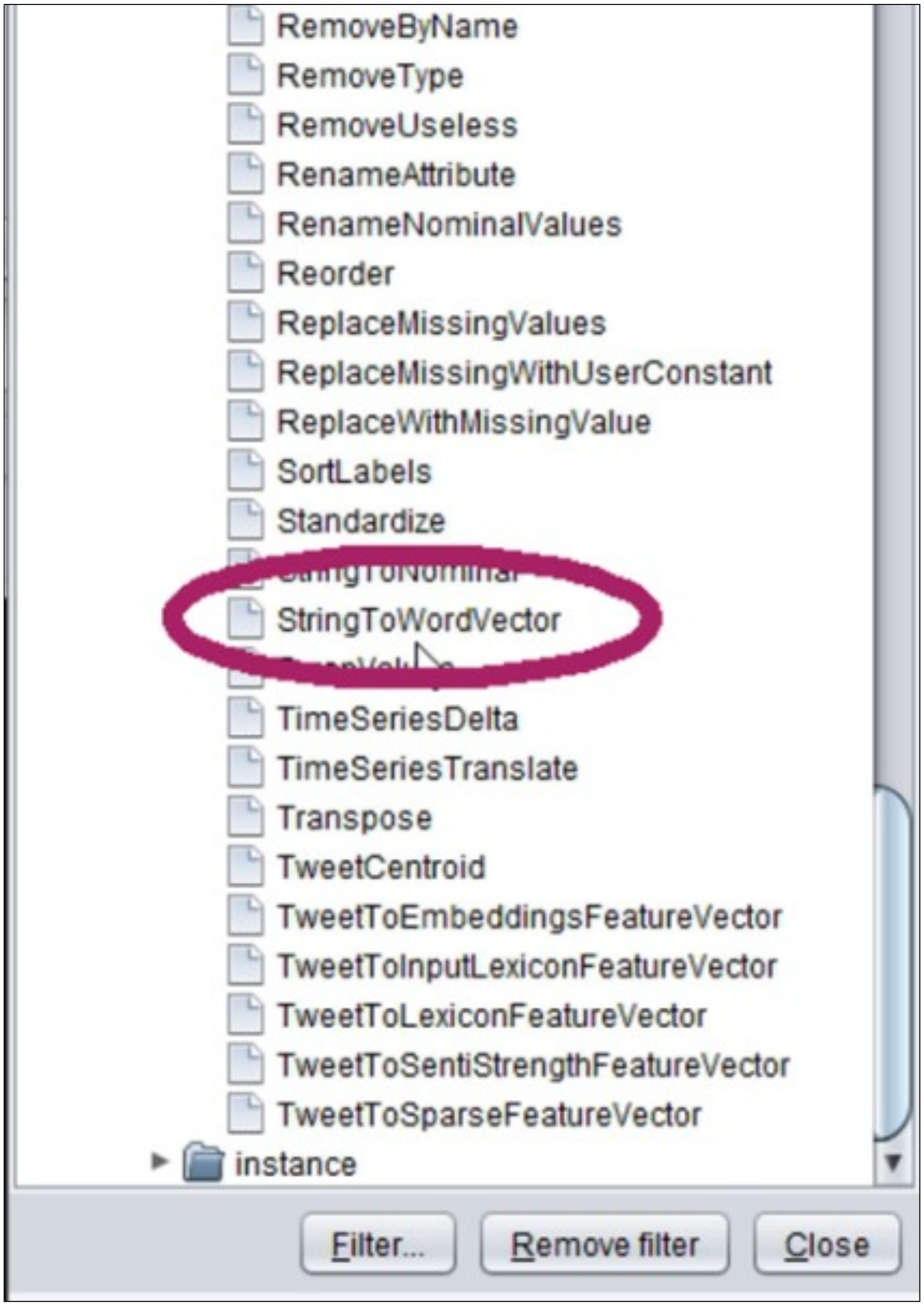
Appendix C.8 Click Choose to open classifiers list



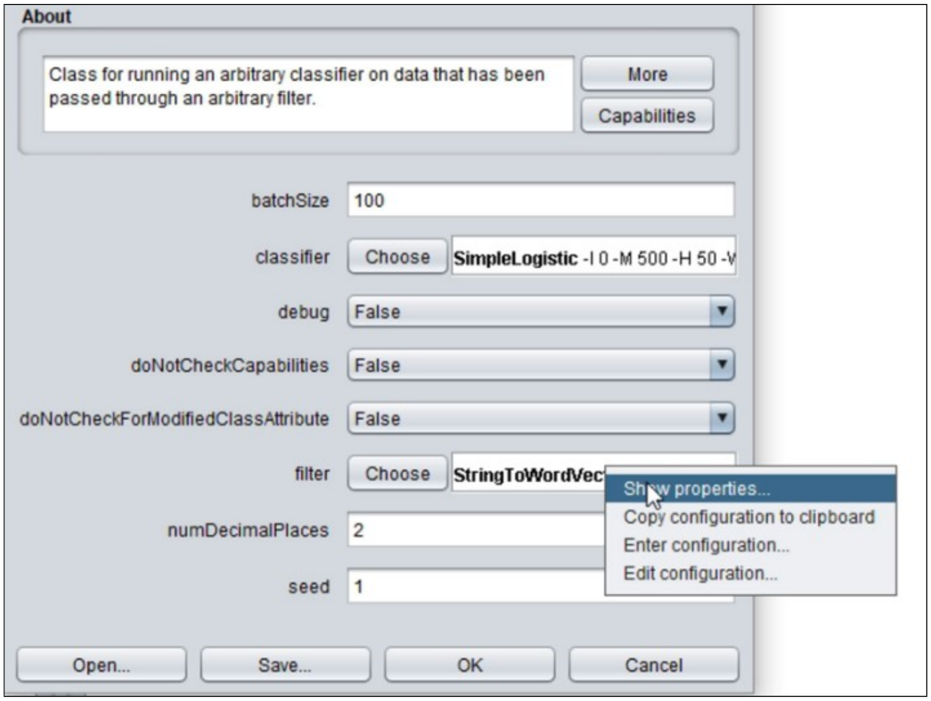
Appendix C.9 Choose a classifier [e.g., simple logistic]



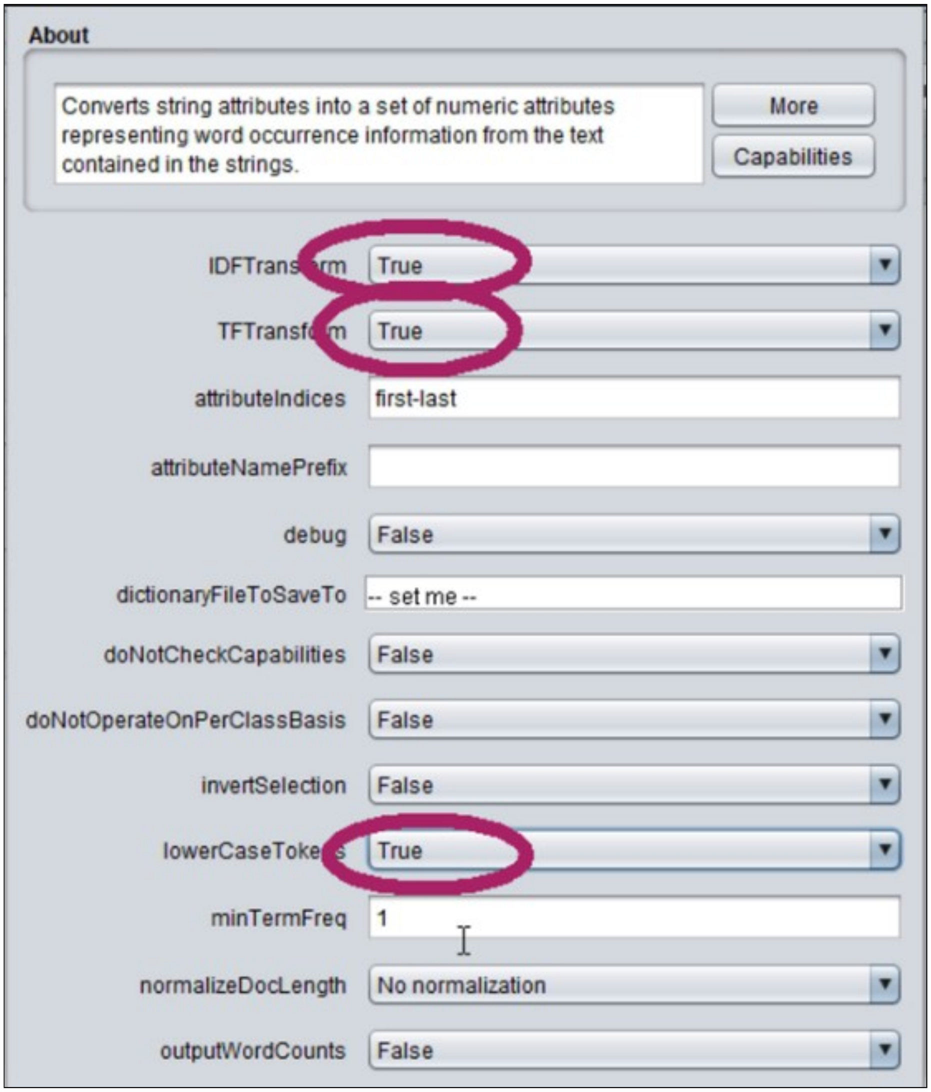
Appendix C.10 Click Choose to open filter options



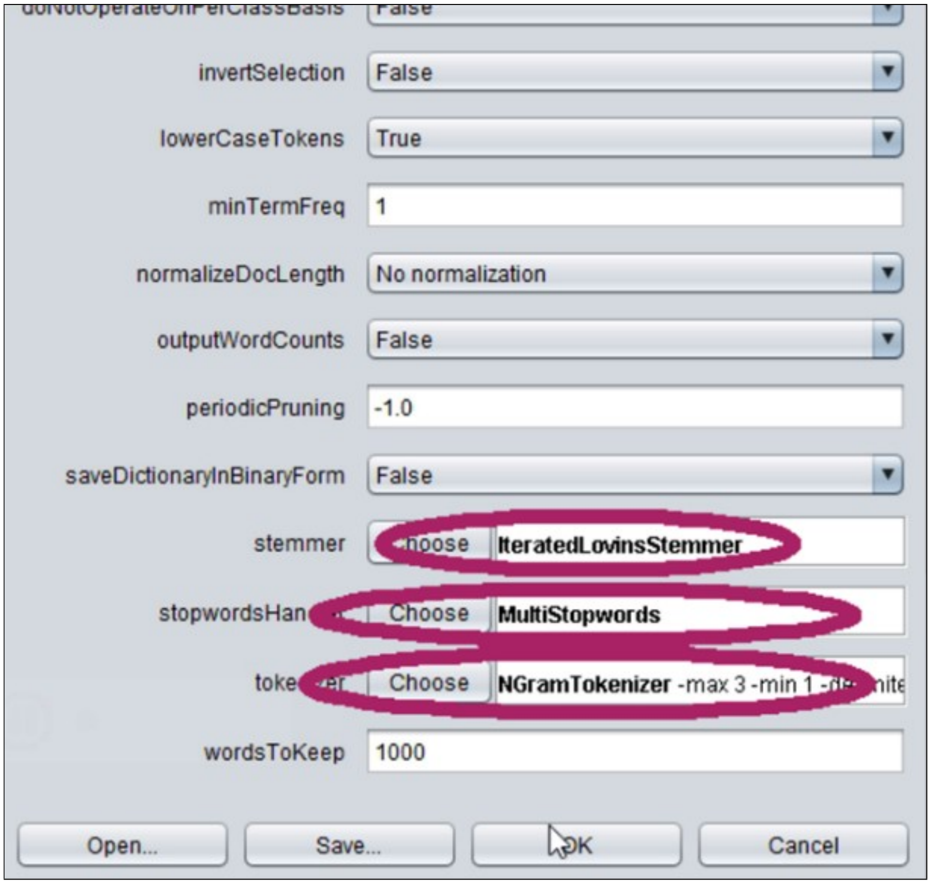
Appendix C.11 Choose StringToWordVector



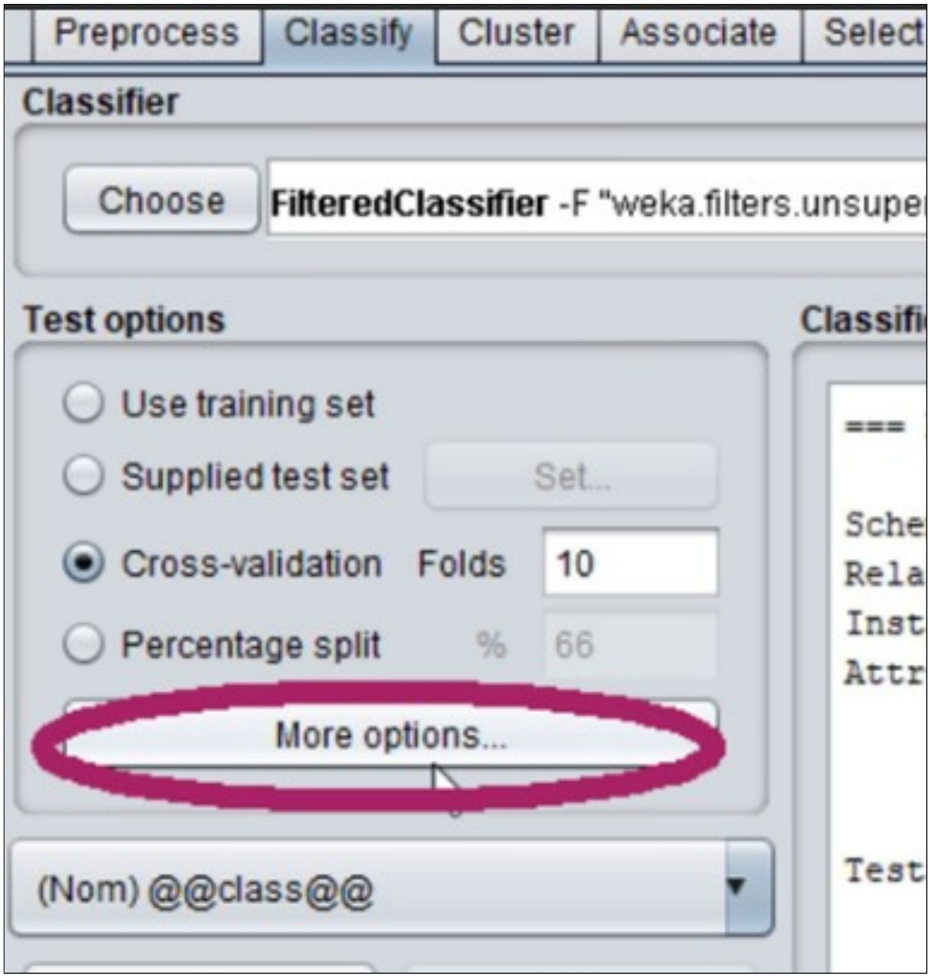
Appendix C.12 Right click on filter name and click show properties



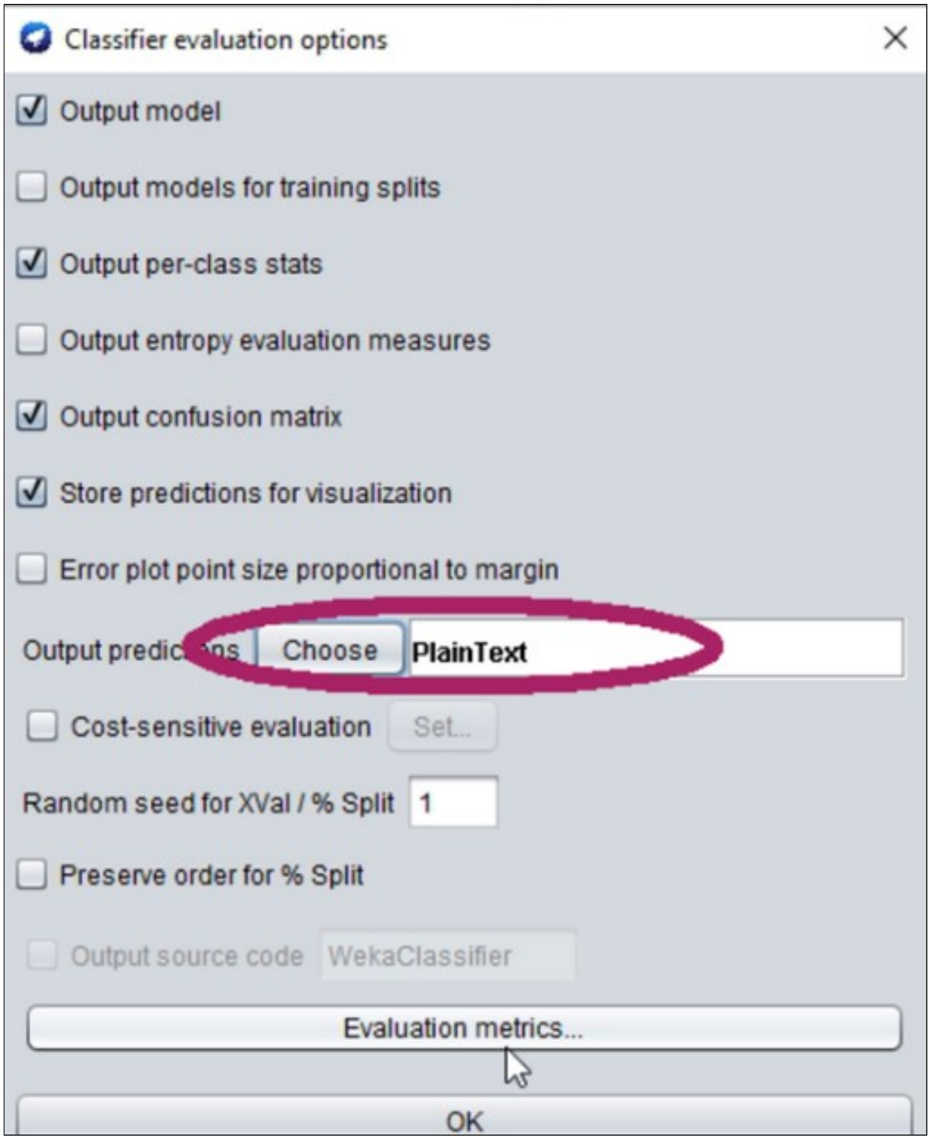
Appendix C.13 Suggested filter settings for text classifications



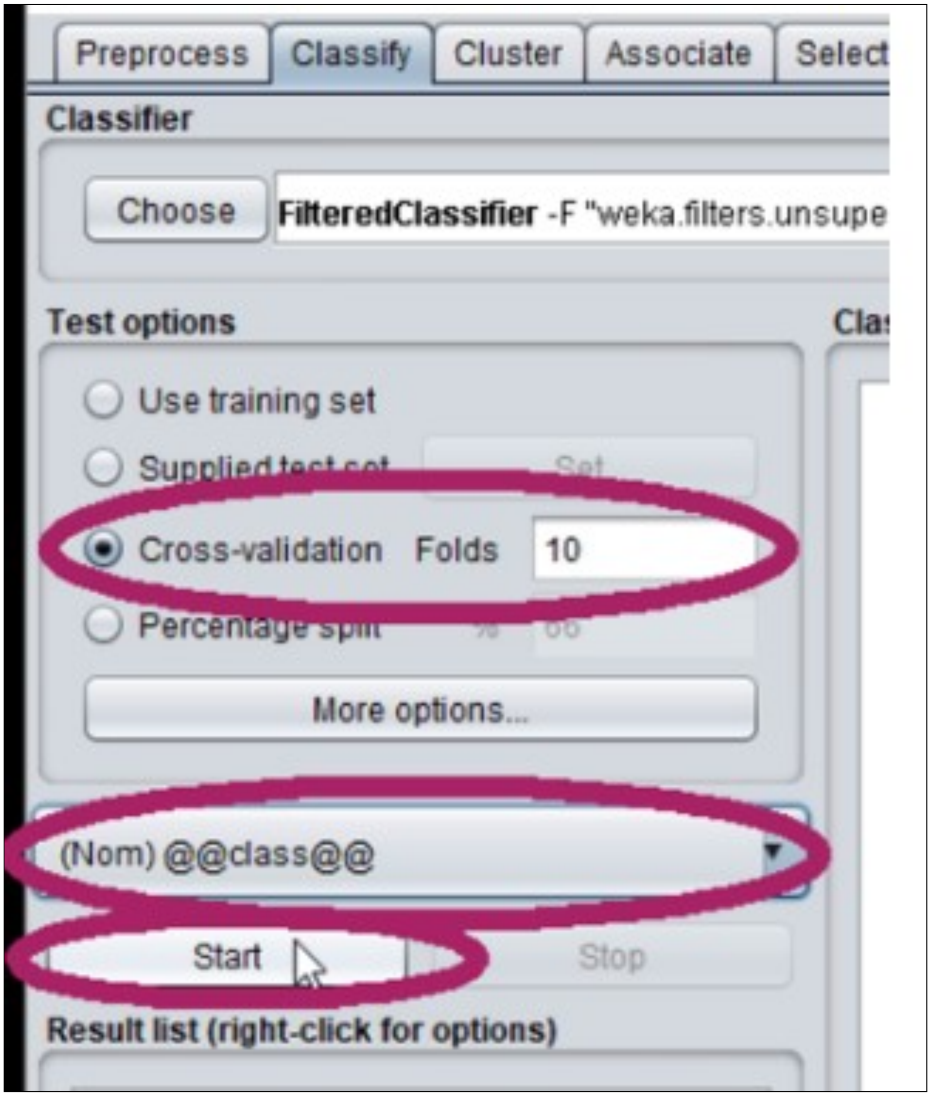
Appendix C.14 Suggested filter settings for text classifications continues



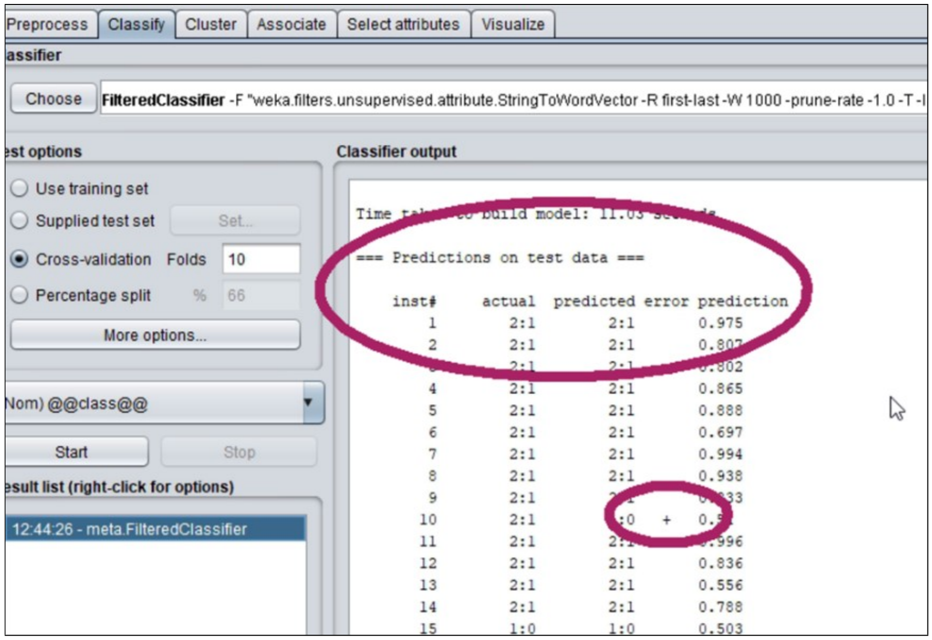
Appendix C.15 From main classify tab, click More options



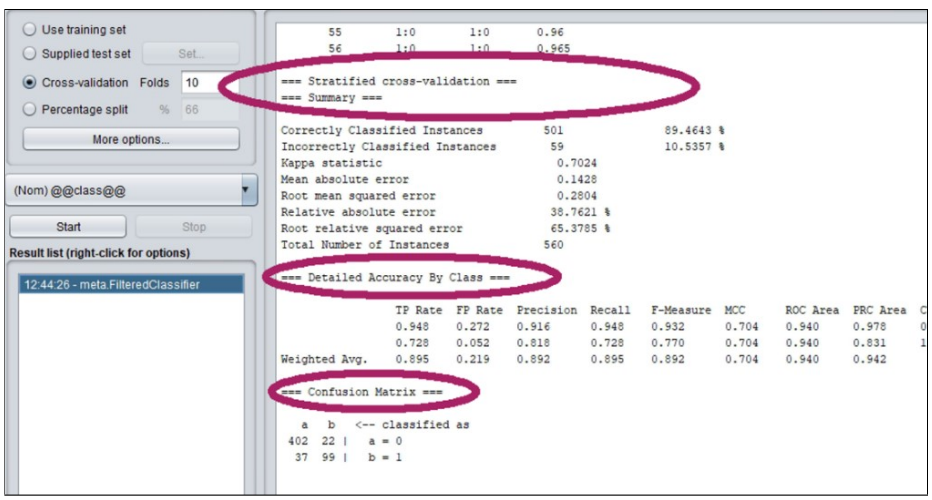
Appendix C.16 Click Choose to open output predictions



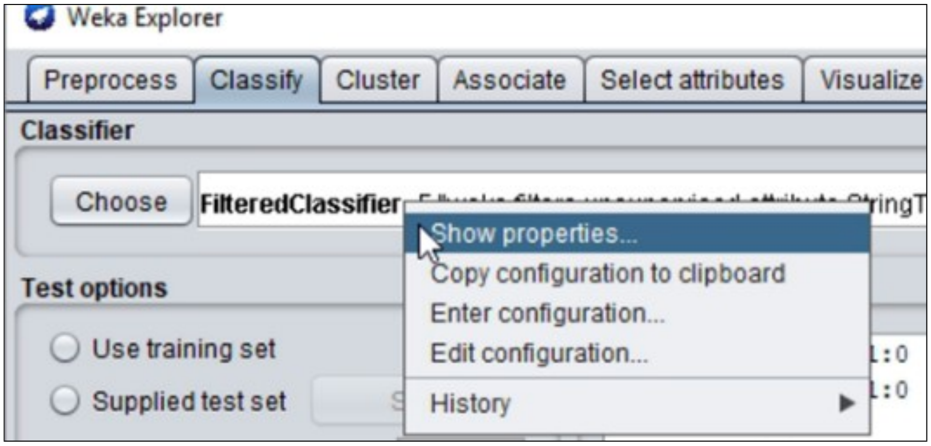
Appendix C.17 Choose cross-validation and classifying attribute



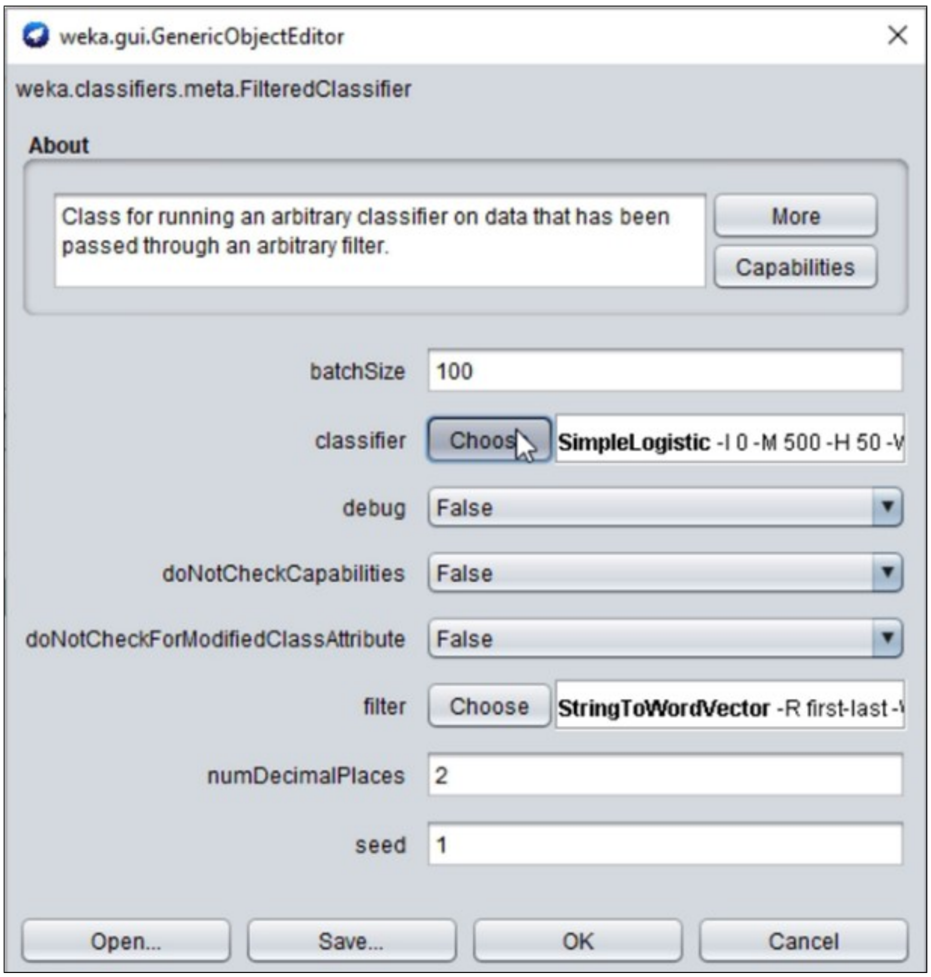
Appendix C.18 Prediction results are shown in the buffer



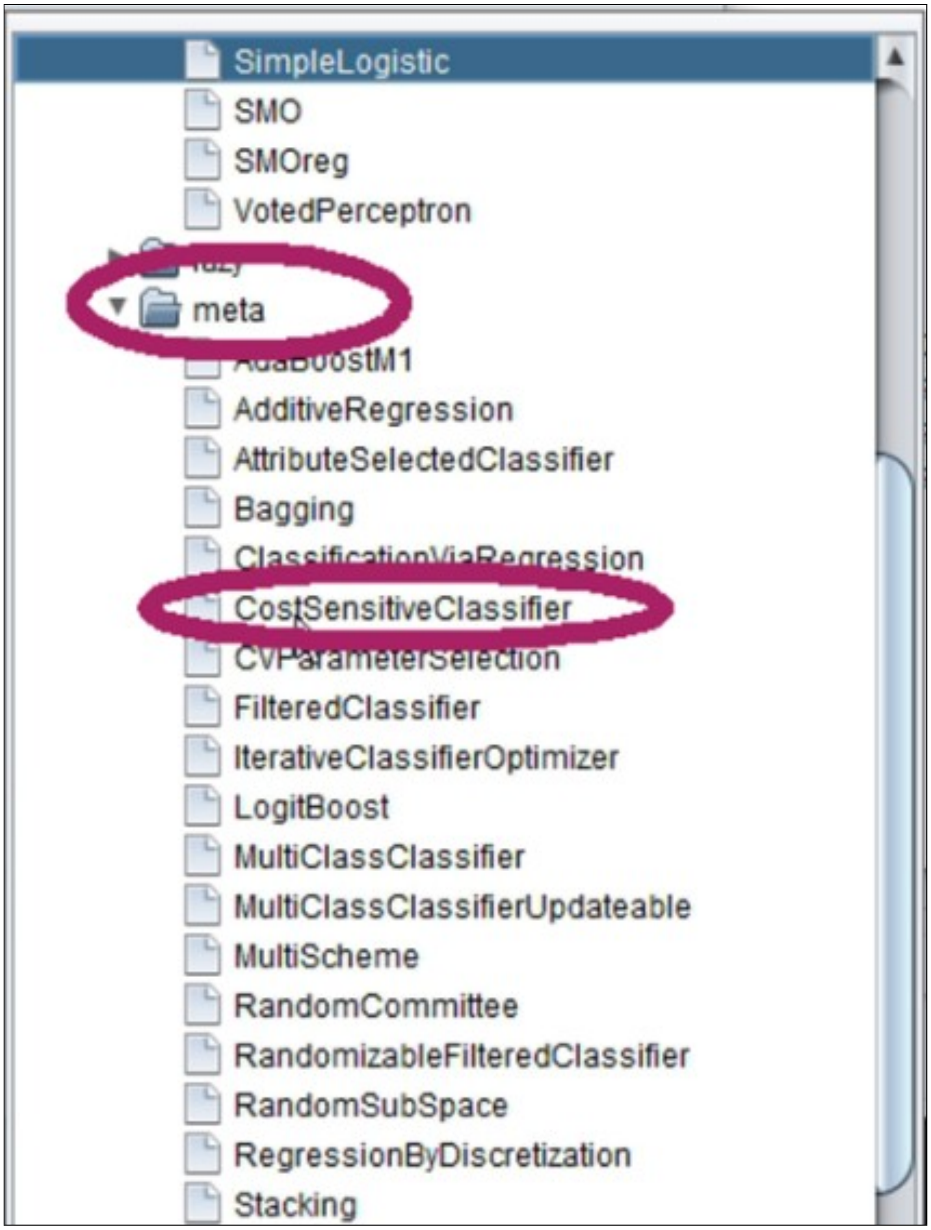
Appendix C.19 Accuracy measures and model performance results are shown at the end of the buffer results



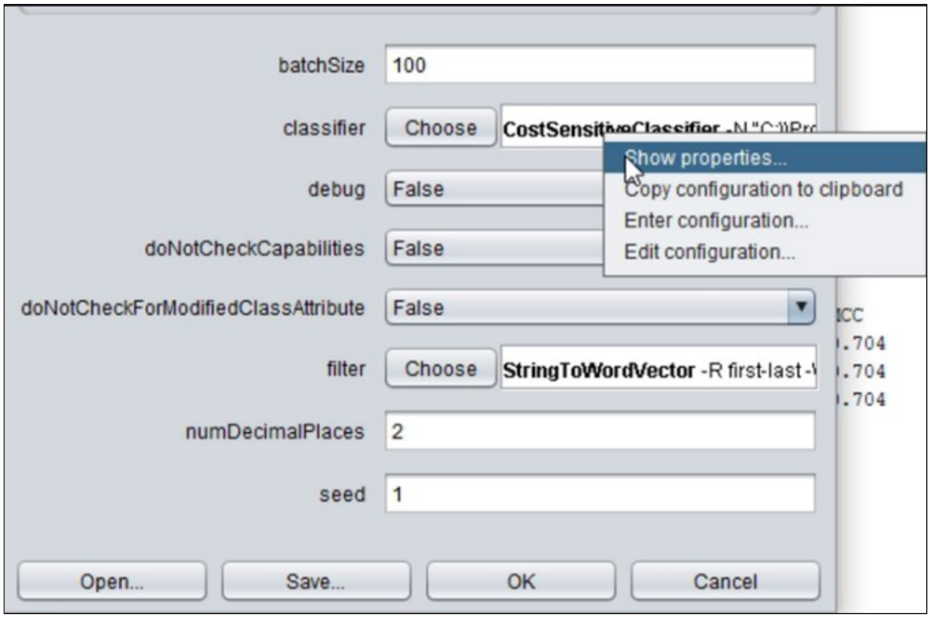
Appendix C.20 Right click on classifier name, then click show properties



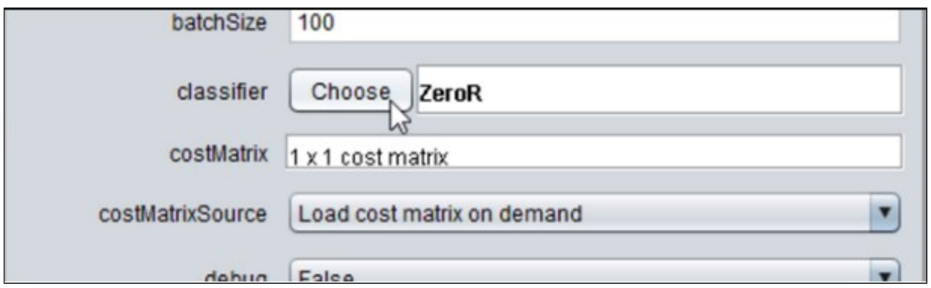
Appendix C.21 Click Choose to open classifier options



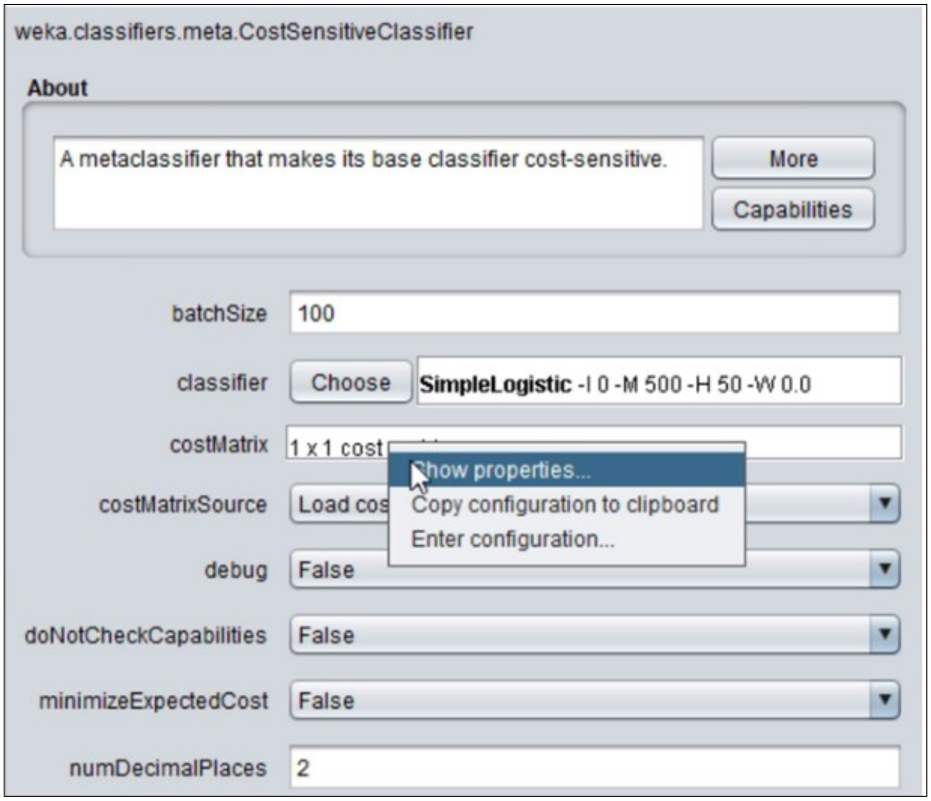
Appendix C.22 Choose Cost Sensitive Classifier



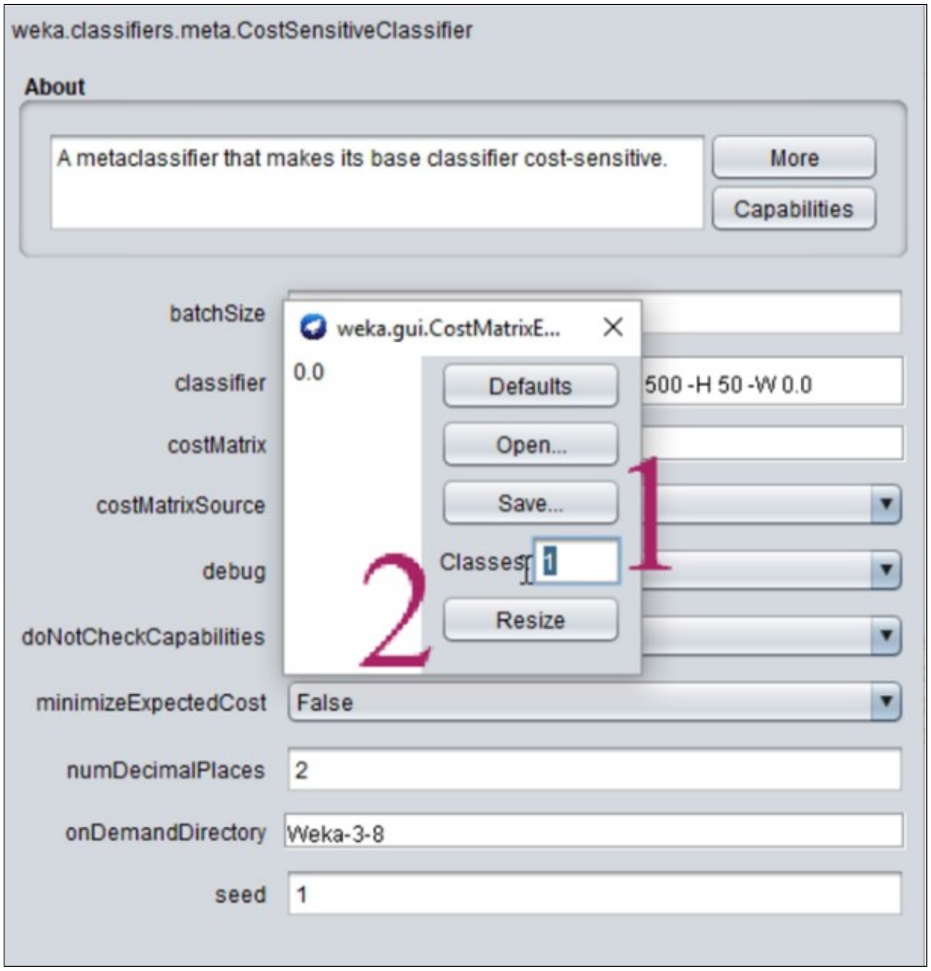
Appendix C.23 Right click to show properties



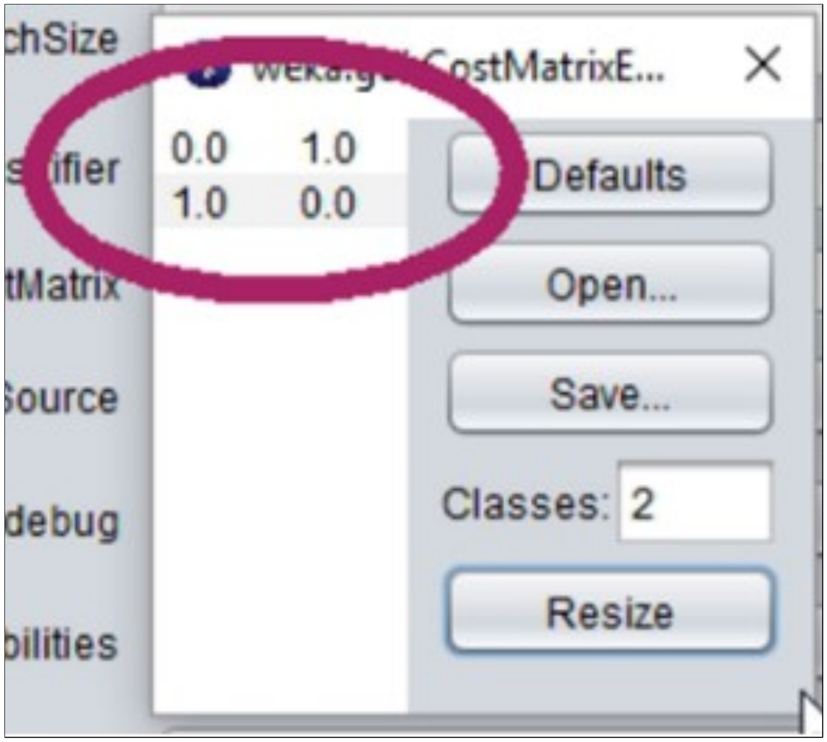
Appendix C.24 Click choose to open classifier options and choose a classifier



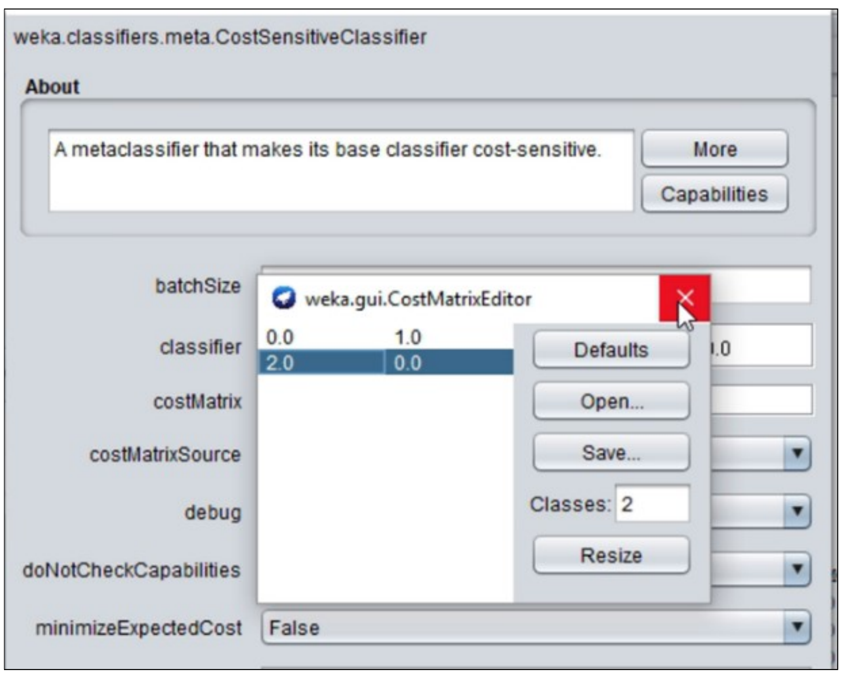
Appendix C.25 Right click on cost matrix to show properties



Appendix C.26 Change classes number according to the study classes, then click resize



Appendix C.27 Cost Matrix now reflects study classes count



Appendix C.28 Adjust penalty as needed. Train the model

D. **JAVA CODES TO EXTRACT TEXT**

Appendix D.1 Codes to extract text from PDF inspired by Dara Yuk original code [129]

```
package abdullah ;
import java . io . File ; import java . io . FileInputStream ; import java . io .
FileOutputStream ; import java . io . FileReader ; import java . io . FileWriter ; import
java . io . BufferedWriter ; import java . io . IOException ; import com. itextpdf . text
. Document ; import com. itextpdf . text . pdf . PdfReader ; import com. itextpdf . text
. pdf . parser . PdfTextExtractor ; import java . awt . Desktop ; import javax . swing .
filechooser . FileNameExtensionFilter ; import javax . swing . JFileChooser ;
/**
  *
  * @author
  */ public      class PDFToTextConverter { public
static void main( String [ ] args ){ selectPDFFiles
();
  }
  //allow  pdf  f  i  l  e  s  selection  for  converting  public  static  void
  selectPDFFiles (){
    JFileChooser chooser = new JFileChooser (); FileNameExtensionFilter f i l t
          e  r  =  new  FileNameExtensionFilter  (”PDF”  ,”  pdf  ”);  chooser  .
          setFileFilter ( f i l t e r ); chooser . setMultiSelectionEnabled ( true );
          int returnVal = chooser . showOpenDialog( null ); String path= F i l e
          P a t h ;
           i f ( returnVal == JFileChooser .APPROVE OPTION) {
                         File [ ] Files=chooser . getSelectedFiles (); System . out . println
                         (”  Please  wait  . . .  ”  )  ;  for  (  int  i  =0;i<Files  .  length  ;  i++){
                         convertPDFToText( Files [ i ] . toString () , path
                        + Files [ i ] . getName (). substring (0 ,
                                          Files [ i ] . getName (). length ()−4)+”. txt ”);
                         }
      System . out . println (” Conversion complete ”);
                           }
             }
public          static          void convertPDFToText( String          src , String       desc ){
    try{
       // create     f i l e     writer
      FileWriter fw=new FileWriter ( desc );
       // create       buffered       writer
      BufferedWriter bw=new BufferedWriter (fw );
       // create       pdf reader
      PdfReader pr=new PdfReader( src ); //get the number of
      pages in the document int pNum=pr . getNumberOfPages ();
      for ( int page=1;page<=pNum; page++){
        String text=PdfTextExtractor . getTextFromPage(pr , page ); bw. write ( text ); bw.
        newLine ();
```

```
      } bw. flush ();
      bw. close ();
    }catch ( Exception e){e . printStackTrace ();}
  }
}
```

Appendix D.2 Codes to extract text from CSV

```
package abdullah ;
import java . io . BufferedReader ; import java . io .
BufferedWriter ; import java . io . File ; import java . io .
FileNotFoundException ; import java . io . FileReader ;
import java . io . FileWriter ; import java . io .
IOException ; import java . util . StringTokenizer ;
/**
  *
  * @author
  */
public      class      Abdullah {
        /**
                 * @param args the command line           arguments
          */  public  static  void  main(  String  [  ]  args  )  throws
        FileNotFoundException , IOException {
                File = new File ( F i l e P a t h ); BufferedReader br = new
                BufferedReader (new FileReader ( f i l e ));
                  String      line ;
                    StringTokenizer       st ;
                int counter = 0; while (( line = br . readLine ()) != null ) { st =
                new StringTokenizer ( line , ”\t ”); i f ( st . countTokens () >=
                3) { st . nextToken ();
                               String name = st . nextToken () + ”. txt ”; BufferedWriter
                               out = new BufferedWriter
                               (new FileWriter (” FilePath” + name ));
                               st . nextToken (); i f    ( st . hasMoreTokens
                               ()) { out . write ( st . nextToken ());
                               } else {
                                           out . write (””);
                               } out . close ();
                               counter++;
            }       // ( st . countTokens()==4) else {
                    System . err . println (” line \t” + counter ); counter++;
            }
    } // while
        } }
```

7. REFERENCES

[1] D. E. Joranson, “et al., trends in medical use and abuse of opioid analgesics,” *Jama*, vol. 283, no. 13, pp. 1710–1714, 2000.

[2] S. E. McCabe, B. T. West, and H. Wechsler, “Trends and college-level characteristics associated with the non-medical use of prescription drugs among us college students from 1993 to 2001,” *Addiction*, vol. 102, no. 3, pp. 455–465, 2007.

[3] R. K. Portenoy, *Chronic opioid therapy in nonmalignant pain*. Journal of pain and symptom management, 1990.

[4] L. Manchikanti, M. B. F. M., and M. H. Ailinani, “Therapeutic use, abuse, and nonmedical use of opioids: a ten-year perspective,” *Pain Physician*, vol. 13, pp. 401–435, 2010.

[5] CDC, “National vital statistics system: Mortality data,” http://www.cdc.gov/nchs/deaths.htm, 2013, accessed: 2018-01-23.

[6] N. Volkow, *American’s Addiction to Opioids: Heroin and Prescription Drug Abuse*, vol. 2014, 2018. [Online]. Available: https://www.drugabuse.gov/about-nida/legislative-activities/testimonyto-congress/2016/americas-addiction-to-opioids-heroin-prescription-drug-abuse

[7] W. M. Compton and N. D. Volkow, “Major increases in opioid analgesic abuse in the united states: concerns and strategies,” *Drug and alcohol dependence*, vol. 81, no. 2, pp. 103–107, 2006.

[8] U. D. Health and H. H. Services, “With chartbook on trends in the health of americans.” *2006 Claitor’s Law Books and Publishing Division*, 2005.

[9] J. Clarke, A. Skoufalos, and R. Scranton, "The american opioid epidemic: Population health implications and potential solutions. report from the national stakeholder panel," *Population health management*, vol. 19, pp. S1–S10, 2016.

[10] L. Manchikanti, "National drug control policy and prescription drug abuse: facts and fallacies," *Pain physician*, vol. 10, no. 3, pp. 399–424, 2007.

[11] L. Manchikanti, E. Whitfield, and F. Pallone, "Evolution of the national all schedules prescription electronic reporting act (nasper): a public law for balancing treatment of pain and drug abuse and diversion," *Pain physician*, vol. 8, no. 4, pp. 335–47, 2005.

[12] L. Manchikanti, "Prescription drug abuse: what is being done to address this new drug epidemic? testimony before the subcommittee on criminal justice, drug policy and human resources," *Pain physician*, vol. 9, no. 4, pp. 287–321, 2006.

[13] L. Manchikanti and A. Singh, "Therapeutic opioids: a ten-year perspective on the complexities and complications of the escalating use, abuse, and nonmedical use of opioids," *Pain physician*, vol. 11, no. 2, pp. 63–88, 2008.

[14] A. M. Gilson and P. G. Kreis, "The burden of the nonmedical use of prescription opioid analgesics," *Pain Medicine*, vol. 10, pp. S89–S100, 2009.

[15] H. G. Birnbaum, "et al., estimated costs of prescription opioid analgesic abuse in the united states in 2001: a societal perspective," *The Clinical journal of pain*, vol. 22, no. 8, pp. 667–76, 2006.

[16] H. G. Birnbaum and et al, “Societal costs of prescription opioid abuse, dependence, and misuse in the united states,” *Pain medicine (Malden, Mass)*, vol. 12, no. 4, pp. 657–67, 2011.

[17] N. National, *Overdose Deaths-Number of Deaths from Prescription Opioid Pain Relievers*, vol. 2015, 2018. [Online]. Available: http://www.drugabuse.gov/related-topics/trends-statistics/overdosedeath-rates

[18] Asam, “Opioid addiction 2016 facts and figures,” *Addiction*, 2016. [Online]. Available: http://www.asam.org/docs/default-source/advocacy/opioidaddiction-disease-facts-figures.pdf

[19] CDC., “Drug overdose deaths in the united states continue to increase in 2015,” 2015. [Online]. Available: https://www.cdc.gov/drugoverdose/epidemic/

[20] C. M. Jones, “Heroin use and heroin use risk behaviors among nonmedical users of prescription opioid pain relievers-united states, 2002-2004 and 2008-2010,” *Drug and alcohol dependence*, vol. 132, no. 1, pp. 95–100, 2013.

[21] R. A. Rudd and et al, “Increases in drug and opioid overdose deaths-united states,” *American Journal of Transplantation*, vol. 16, no. 4, pp. 1323–1327, 2016.

[22] Cdc., “Opioid painkiller prescribing,” p. 1, 2014. [Online]. Available: https://www.cdc.gov/vitalsigns/opioid-prescribing/

[23] CDC, “Prescription painkiller overdoses,” vol. 2018, p. 1, 2013. [Online]. Available: https://www.cdc.gov/vitalsigns/prescriptionpainkilleroverdoses/index.html

[24] Samhsa., “Key substance use and mental health indicators in the united states: Results from the 2015 national survey on drug use and health,” 2016. [Online]. Available: https://www.samhsa.gov/data/sites/default/files/NSDUH-FFR12015/NSDUH-FFR1-2015/NSDUH-FFR1-2015.pdf

[25] Abuse, “N. i.o.d. prescription and over-the-counter medications,” p. 1, 2015. [Online]. Available: https://www.drugabuse.gov/publications/drugfacts/prescription-over-countermedications

[26] R. J. Fortuna and et al., “Prescribing of controlled medications to adolescents and young adults in the united states,” *Pediatrics*, vol. 126, no. 6, pp. 1108–1116, 2010.

[27] J. Duwve *et al.*, “Report on the toll of opioid use in indiana and marion county,” vol. 2016, 2018. [Online]. Available: https://www.inphilanthropy.org/sites/default/files/Richard

[28] C. A. Harle and et al., “Decision support for chronic pain care: how do primary care physicians decide when to prescribe opioids? a qualitative study,” *BMC family practice*, vol. 16, no. 1, p. 48, 2015.

[29] J. H. Chen and et al., “Distribution of opioids by different types of medicare prescribers,” *JAMA internal medicine*, vol. 176, no. 2, pp. 259–261, 2016.

[30] P. Bendtsen and et al, “What are the qualities of dilemmas experienced when prescribing opioids in general practice?” *Pain*, vol. 82, no. 1, pp. 89–96, 1999.

[31] A. A. Bergman and et al., “Contrasting tensions between patients and pcps in chronic pain management: a qualitative study,” *Pain Medicine*, vol. 14, no. 11, pp. 1689–1697, 2013.

[32] R. R. Leverence and et al., “Chronic non-cancer pain: a siren for primary carea report from the primary care multiethnic network (prime net),” *The Journal of the American Board of Family Medicine*, vol. 24, no. 5, pp. 551–561, 2011.

[33] M. S. Matthias and et al., “The patient-provider relationship in chronic pain care: providers' perspectives,” *Pain Medicine*, vol. 11, no. 11, pp. 1688–1697, 2010.

[34] A. Spitz and et al., “Primary care providers' perspective on prescribing opioids to older adults with chronic non-cancer pain: a qualitative study,” *BMC geriatrics*, vol. 11, no. 1, p. 35, 2011.

[35] S. F. a. e. a. Butler, “Validation of a screener and opioid assessment measure for patients with chronic pain,” *Pain*, vol. 112, no. 1-2, pp. 65–75, 2004.

[36] H. Akbik and et al., “Validation and clinical application of the screener and opioid assessment for patients with pain (soapp),” *Journal of Pain & Symptom Management*, vol. 32, no. 3, pp. 287–93, 2006.

[37] S. F. Butler and et al., “Validation of the revised screener and opioid assessment for patients with pain (soapp-r),” *Journal of Pain*, vol. 9, no. 4, pp. 360–72, 2008.

[38] Clin J Pain, “Cross validation of the current opioid misuse measure to monitor chronic pain patients on opioid therapy,” *Clinical Journal of Pain*, vol. 26, no. 9, pp. 770–6, 2010.

[39] S. M. Wu and et al., "The addiction behaviors checklist: validation of a new clinician-based measure of inappropriate opioid use in chronic pain," *Journal of pain and symptom management*, vol. 32, no. 4, pp. 342–351, 2006.

[40] T. Jones and A. S. D. Passik, "comparison of methods of administering the opioid risk tool," *Journal of Opioid Management*, vol. 7, no. 5, pp. 347–51, 2011.

[41] S. F. Butler and et al., "Electronic opioid risk assessment program for chronic pain patients: Barriers and benefits of implementation," *Pain Practice*, vol. 14, no. 3, pp. E98–E105, 2014.

[42] Inflexxion, "Soapp long form vs. soapp short form: Tradeoff in accuracy," 2005. [Online]. Available: https://www.painedu.com/soapp/SOAPP tradeoff.pdf

[43] T. Jones and et al., "A comparison of various risk screening methods in predicting discharge from opioid treatment," *Clinical Journal of Pain*, vol. 28, no. 2, pp. 93–100, 2012.

[44] T. M. Moore, , and et al., "A comparison of common screening methods for predicting aberrant drug-related behavior among patients receiving opioids for chronic pain management," *Pain Medicine*, vol. 10, no. 8, pp. 1426–33, 2009.

[45] R. West and J. Brown, *Theory of Addiction*, 2nd ed.-Blackwell: Wiley, 2013.

[46] W.S. abuse, p. 1, 2016. [Online]. Available: http://www.who.int/topics/substance abuse/en/

[47] "Triwest healthcare alliance. substance abuse vs substance dependence," 2016. [Online]. Available: http://www.triwest.com/en/behavioral-health/resourcecenter/addiction-recovery/substance-abuse/more/

[48] "Highlights of changes from dsm-iv-tr to dsm-5," 2013. [Online]. Available: http://www.dsm5.org/documents/changes

[49] C. Tools, "Dsm 5 criteria for substance use disorder," 2013. [Online]. Available: http://www.buppractice.com/node/12351

[50] N. I. Abuse, "Aberrant drug taking behaviors," vol. 4, 2016. [Online]. Available: https://www.drugabuse.gov/sites/default/files/files/AberrantDrugTakingBehaviors.pdf

[51] J. Jaffe, C. Knapp, and D. Ciraulo, "Opiates: clinical aspects substance abuse: a comprehensive text," *Williams and Wilkens*, pp. 158–66, 1997.

[52] M. J. Cohen and et al., "Ethical perspectives: Opioid treatment of chronic pain in the context of addiction," *Clinical Journal of Pain*, vol. 18, no. 4, pp. S99–S107, 2002.

[53] E. Michna and et al., "Predicting aberrant drug behavior in patients treated for chronic pain: importance of abuse history," *Journal of pain and symptom management*, vol. 28, no. 3, pp. 250–258, 2004.

[54] Drugabuse.gov, "Aberrant drug taking behaviors information sheet," vol. 1, 2018. [Online].Available: https://www.drugabuse.gov/sites/default/files/files/AberrantDrugTakingBehaviors.pdf

[55] T. J. Ives and et al., "Predictors of opioid misuse in patients with chronic pain: a prospective cohort study," *BMC Health Services Research*, vol. 6, p. 46, 2006.

[56] M. J. Edlund and et al., "Risk factors for clinically recognized opioid abuse and dependence among veterans using opioids for chronic non-cancer pain," *Pain*, vol. 129, no. 3, pp. 355–62, 2007.

[57] A. G. White and et al., "Analytic models to identify patients at risk for prescription opioid abuse," *American Journal of Managed Care*, vol. 15, no. 12, pp. 897–906, 2009.

[58] J. B. Rice and et al., "A model to identify patients at risk for prescription opioid abuse, dependence, and misuse," *Pain Medicine*, vol. 13, no. 9, pp. 1162–73, 2012.

[59] M. Lewis, C. M. Herndon, and J. T. Chibnall, "Patient aberrant drug taking behaviors in a large family medicine residency program: a retrospective chart review of screening practices, incidence, and predictors," *Journal of Opioid Management*, vol. 10, no. 3, pp. 169–75, 2014.

[60] J. A. Turner and et al., "Chronic opioid therapy urine drug testing in primary care: prevalence and predictors of aberrant results," *Journal of General Internal Medicine*, vol. 29, no. 12, pp. 1663–71, 2014.

[61] T. R. Hylan and et al., "Automated prediction of risk for problem opioid use in a primary care setting," *Journal of Pain*, vol. 16, no. 4, pp. 380–7, 2015.

[62] P. D. Allison, *Logistic regression using SAS:Theory and application SAS Institute*, S. Institute, Ed., 2012.

[63] E. L. Zale and et al., "Tobacco smoking, nicotine dependence, and patterns of prescription opioid misuse: Results from a nationally representative sample," *Nicotine & Tobacco Research*, vol. 17, no. 9, pp. 1096–103, 2015.

[64] J. Wilson and A. Bock, “The benefit of using both claims data and electronic medical record data in health care analysis,” 2012. [Online]. Available: https://www.optum.com/content/dam/optum/resources/whitePapers/Benefitsof-using-both-claims-and-EMR-data-in-HC-analysis-WhitePaper-ACS.pdf

[65] E. F. Kern and et al., “Failure of icd?9?cm codes to identify patients with comorbid chronic kidney disease in diabetes,” *Health services research*, vol. 41, no. 2, pp. 564–580, 2006.

[66] M. L. Hansen, P. W. Gunn, and D. C. Kaelber, “Underdiagnosis of hypertension in children and adolescents,” *Jama*, vol. 298, no. 8, pp. 874–879, 2007.

[67] S. F. Butler and et al., “Cross-validation of a screener to predict opioid misuse in chronic pain patients (soapp-r),” *Journal of addiction medicine*, vol. 3, no. 2, p. 66, 2009.

[68] Inflexxion,“Soapp,” 2015. [Online].Available: https://www.inflexxion.com/soapp-comm/

[69] J. S. Knisely and et al., “Prescription opioid misuse index: a brief questionnaire to assess misuse,” *Journal of Substance Abuse Treatment*, vol. 35, no. 4, pp. 380–6, 2008.

[70] T. Jones and et al., “Further validation of an opioid risk assessment tool: the brief risk interview,” *Journal of Opioid Management*, vol. 10, no. 5, pp. 353–64, 2014.

[71] Zdnet, Within two years, 2013. [Online]. Available: http://www.zdnet.com/article/within-two-years-80-of-all-medical-datawill-be-unstructured/

[72] S. M. Smith and et al., “Classification and definition of misuse, abuse, and related events in clinical trials: Acttion systematic review and recommendations,” *PAIN*, vol. 154, no. 11, pp. 2287–2296, 2013.

[73] T. P. Stratton, L. Palombi, H. Blue, and M. E. Schneiderhan, “Ethical dimensions of the prescription opioid abuse crisis,” *American Journal of HealthSystem Pharmacy*, vol. 75, no. 15, pp. 1145–1150, 2018.

[74] C. Gounder, “Who is responsible for the pain-pill epidemic,” *The New Yorker*, 2013.

[75] B. Meier, *Pain killer: A” wonder” drug’s trail of addiction and death*. Rodale, 2003.

[76] G. Baucus, “Baucus Grassley opioid investigation letter to American pain foundation,” https://www.finance.senate.gov/download/baucus-grassleyopioid-investigation-letter-to-american-pain-foundation, May 2012, accessed: 2018-08-20.

[77] G. Kounang, “Senate report says patient advocacy groups get kickbacks from opioid-manufacturers,” https://www.cnn.com/2018/02/12/health/senatereport-opioid-manufacturers-donations/index.html , February 2018, accessed: 2018-08-20.

[78] Ornstein, “American pain foundation shuts down as senators launch investigation of prescription narcotics,” https://www.propublica.org/article/senatepanel-investigates-drug-company-ties-to-pain-groups, May 2012, accessed: 2018-08-20.

[79] R. C. Dart, H. L. Surratt, T. J. Cicero, M. W. Parrino, S. G. Severtson, B. Bucher-Bartelson, and J. L. Green, "Trends in opioid analgesic abuse and mortality in the united states," *New England Journal of Medicine*, vol. 372, no. 3, pp. 241–248, 2015.

[80] T. D. Saha, B. T. Kerridge, R. B. Goldstein, S. P. Chou, H. Zhang, J. Jung, R. P. Pickering, W. J. Ruan, S. M. Smith, B. Huang *et al.*, "Nonmedical prescription opioid use and dsm-5 nonmedical prescription opioid use disorder in the united states," *The Journal of clinical psychiatry*, vol. 77, no. 6, p. 772, 2016.

[81] S. M. Frenk, K. S. Porter, and L. Paulozzi, *Prescription opioid analgesic use among adults: United States, 1999-2012*. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, 2015, no. 2015.

[82] A. Kolodny, D. T. Courtwright, C. S. Hwang, P. Kreiner, J. L. Eadie, T. W. Clark, and G. C. Alexander, "The prescription opioid and heroin crisis: a public health approach to an epidemic of addiction," *Annual review of public health*, vol. 36, pp. 559–574, 2015.

[83] Cdc., "Increases in drug and opioid-involved overdose deaths united states, 20102015," 2016. [Online]. Available: https://www.cdc.gov/mmwr/volumes/65/wr/mm655051e1.htm

[84] national institute of drug abuse, "Indiana opioid summary," https://www.drugabuse.gov/drugs-abuse/opioids/opioid-summaries-bystate/indiana-opioid-summary, February 2018, accessed: 2018-08-21.

[85] A. H. Alzeer, J. Jones, and M. J. Bair, "Review of factors, methods, and outcome definition in designing opioid abuse predictive models," *Pain Medicine*, vol. 19, no. 5, pp. 997–1009, 2017.

[86] U. Raja and et al., "Text mining in healthcare. applications and opportunities," *Journal of healthcare information management: JHIM*, vol. 22, no. 3, pp. 52–56, 2008.

[87] R. Kukafka and et al., "Human and automated coding of rehabilitation discharge summaries according to the international classification of functioning, disability, and health," *Journal of the American Medical Informatics Association*, vol. 13, no. 5, pp. 508–515, 2006.

[88] S. G. Peters and J. D. Buntrock, "Big data and the electronic health record," *Journal of Ambulatory Care Management*, vol. 37, no. 3, pp. 206–10, 2014.

[89] L. Pereira *et al.*, *Using text mining to diagnose and classify epilepsy in children. in e-Health Networking, Applications & Services (Healthcom), 2013 IEEE 15th International Conference on. 2013*. IEEE.

[90] D. S. Carrell and et al., "Using natural language processing to identify problem usage of prescription opioids," *International journal of medical informatics*, vol. 84, no. 12, pp. 1057–1064, 2015.

[91] R. E. Palmer and et al., "The prevalence of problem opioid use in patients receiving chronic opioid therapy: computer-assisted review of electronic health record clinical notes," *Pain*, vol. 156, no. 7, pp. 1208–14, 2015.

[92] I. Hleath, "About indiana university health," 2013. [Online]. Available: http://iuhealth.org/images/glo-new-doc/IUH-FactSheet-07-22-131(2).pdf

[93] A. Ghaffarinejad and M. Kerdegary, "Relationship of opioid dependence and positive and negative symptoms in schizophrenic patients," *Addiction & health*, vol. 1, no. 2, p. 69, 2009.

[94] Regenstrief, "ndepth," 2016. [Online]. Available: http://www.regenstrief.org/resources/ndepth/

[95] Regenstrief_institute, "Regenstrief data core," 2018. [Online]. Available: https://www.indianactsi.org/translational-informatics/data-core/

[96] H. Harkema and et al., "Context: an algorithm for determining negation, experiencer, and temporal status from clinical reports," *Journal of biomedical informatics*, vol. 42, no. 5, pp. 839–851, 2009.

[97] Cornell, "42 cfr part 2 - confidentiality of substance use disorder patient records," https://www.law.cornell.edu/cfr/text/42/part-2, February 2017, accessed: 2018-09-15.

[98] CDC., "U. s. prescribing rate maps," 2017. [Onlin Available: https://www.cdc.gov/drugoverdose/maps/rxrate-maps.html

[99] A. Goodnough, *Opioid Prescriptions Fall After 2010 Peak, C.D.C. Report Finds, in The New York Times*, 2017.

[100] K. Newman, O. Prescriptions and A. D. D. in US News, Eds., 2018.

[101] A. S. Bohnert and et al., "Association between opioid prescribing patterns and opioid overdose-related deaths," *Jama*, vol. 305, no. 13, pp. 1315–1321, 2011.

[102] K. M. Dunn and et al., "Opioid prescriptions for chronic pain and overdose: A cohort study," *Annals of Internal Medicine*, vol. 152, no. 2, pp. 85–92, 2010.

[103] W. M. Compton, C. M. Jones, and G. T. Baldwin, "Relationship between nonmedical prescription-opioid use and heroin use," *New England Journal of Medicine*, vol. 374, no. 2, pp. 154–163, 2016.

[104] M. Okamoto *et al.*, "An asymptotic expansion for the distribution of the linear discriminant function," *The Annals of Mathematical Statistics*, vol. 34, no. 4, pp. 1286–1301, 1963.

[105] J. Beck, "Class balancing in weka," 2018. [Onlin Available: https://www.youtube.com/watch?v=V9PNyx5-kxM

[106] O. Kwon and J. M. Sim, "Effects of data set features on the performances of classification algorithms," *Expert Systems with Applications*, vol. 40, no. 5, pp. 1847–1857, 2013.

[107] V. Lo´pez, A. Fern´andez, S. Garc´ıa, V. Palade, and F. Herrera, "An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics," *Information Sciences*, vol. 250, pp. 113–141, 2013.

[108] Y. Sun, M. S. Kamel, and Y. Wang, "Boosting for learning multiple classes with imbalanced class distribution," in *null*. IEEE, 2006, pp. 592–602.

[109] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge & Data Engineering*, no. 9, pp. 1263–1284, 2008.

[110] G. E. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD explorations newsletter*, vol. 6, no. 1, pp. 20–29, 2004.

[111] B. Zadrozny and C. Elkan, “Learning and making decisions when costs and probabilities are both unknown,” in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2001, pp. 204–213.

[112] P. Domingos, “Metacost: A general method for making classifiers costsensitive,” in *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1999, pp. 155–164.

[113] S. Lomax, “How to know that our dataset is imbalance?” 2017. [Online]. Available:
https://www.researchgate.net/post/How _to know that _our dataset is imbalance

[114] Y. Sun, A. K. Wong, and M. S. Kamel, “Classification of imbalanced data: A review,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, no. 04, pp. 687–719, 2009.

[115] Z.-H. Zhou and X.-Y. Liu, “Training cost-sensitive neural networks with methods addressing the class imbalance problem,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 63–77, 2006.

[116] “.model file,” http://filext.com/file-extension/MODEL, accessed: 2018-08-15.

[117] R. Caruana and A. Niculescu-Mizil, “An empirical comparison of supervised learning algorithms,” in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 161–168.

[118] R. Couronn´e, P. Probst, and A.-L. Boulesteix, "Random forest versus logistic regression: a large-scale benchmark experiment," *BMC bioinformatics*, vol. 19, no. 1, p. 270, 2018.

[119] T. Pranckeviˇcius and V. Marcinkeviˇcius, "Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification," *Baltic Journal of Modern Computing*, vol. 5, no. 2, p. 221, 2017.

[120] C. Perlich, "What are the advantages of logistic regression over decision trees? are there any cases where it's better to use logistic regression instead of decision trees?" https://www.quora.com/What-are-the-advantages-oflogistic-regression-over-decision-trees-Are-there-any-cases-where-its-better-touse-logistic-regression-instead-of-decision-trees/answer/Claudia-Perlich, Jun 2017, accessed: 2018-08-16.

[121] L. M. Schwartz, S. Woloshin, H. C. Sox, B. Fischhoff, and H. G. Welch, "Us women's attitudes to false positive mammography results and detection of ductal carcinoma in situ: cross sectional survey," *Bmj*, vol. 320, no. 7250, pp. 1635–1640, 2000.

[122] J. G. Elmore, D. L. Miglioretti, L. M. Reisch, M. B. Barton, W. Kreuter, C. L. Christiansen, and S. W. Fletcher, "Screening mammograms by community radiologists: variability in false-positive rates," *Journal of the National Cancer Institute*, vol. 94, no. 18, pp. 1373–1380, 2002.

[123] K. F. Foundation, "Opioid overdose death rates and all drug overdose death rates per 100,000 population (age-adjusted),"

https://www.kff.org/other/stateindicator/opioid-overdose-death-rates/, 2018, accessed: 2018-08-29.

[124] S. Redmore, "Machine learning vs. natural language processing," 2013.

[125] R. Zhang, S. Ma, L. Shanahan, J. Munroe, S. Horn, and S. Speedie, "Automatic methods to extract new york heart association classification from clinical notes," in *Bioinformatics and Biomedicine (BIBM), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1296–1299.

[126] C.-H. Lee, C.-H. Wu, and H.-C. Yang, "Text mining of clinical records for cancer diagnosis," in *Innovative Computing, Information and Control, 2007. ICICIC'07. Second International Conference on*. IEEE, 2007, pp. 172–172.

[127] X. Dong, L. Qian, Y. Guan, L. Huang, Q. Yu, and J. Yang, "A multiclass classification method based on deep learning for named entity recognition in electronic medical records," in *Scientific Data Summit (NYSDS), 2016 New York*. IEEE, 2016, pp. 1–10.

[128] Z. Wei, Z. X. Ju, X. Chun, J. Hua, and P. Jin, "An automatic electronic nursing records analysis system based on the text classification and machine learning," in *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2013 5th International Conference on*, vol. 2. IEEE, 2013, pp. 494–498.

[129] D. Yuk, "Pdf to text converter," http://java.worldbestlearningcenter.com/2013/06/pdf-to-text-converter.html, June 2013, accessed: 2018-08-05.

CURRICULUM VITAE

Abdullah Hamad Alzeer

EDUCATION

Master of Science in Health Informatics at Northern Kentucky University, May 2012.

Bachelor of Science in Pharmaceutical science at King Saud University, February 2009.

EMPLOYMENT

Undergraduate Instructor, Department of Health Information Management, Indiana University, January 2017 - 2018.

Teaching Assistant, Department of Clinical Pharmacy, King Saud University, June 2011 - 2018.

Pharmacist, Pharmacy Department, King Fahad Medical City, Riyadh, Saudi Arabia, May 2009 - May 2010.

PUBLICATION

Alzeer, A., Patel, J., Dixon, B., Bair, M. and Jones, J., 2018, June. A Comparison Between Two Approaches to Identify Opioid Use Problems: ICD-9 vs. Text-Mining Approach. In 2018 IEEE International Conference on Healthcare Informatics (ICHI) (pp. 455-456). IEEE.

Alzeer, A.H., Jones, J. and Bair, M.J., 2017. Review of factors, methods, and outcome definition in designing opioid abuse predictive models. Pain Medicine, 19(5), pp.997-1009.

Dixon, B.E., Alzeer, A.H., Phillips, E.O.K. and Marrero, D.G., 2016. Integration of provider, pharmacy, and patient-reported data to improve medication adherence for type 2 diabetes: a controlled before-after pilot study. JMIR medical informatics, 4(1).

PROFESSIONAL PRESENTATIONS

Alzeer, Abdullah H. Identify Opioid Use Problem: Text Mining Approach IEEE ICHI 2018. New York, NY. June 2018.

AWARDS

Best Saudi Students Club President in US by Saudi Arabian Culture Mission. Washington, DC. October 2017.