Rough-set based learning methods: A case study to assess the relationship between the clinical delivery of cannabinoid medicine for anxiety, depression, sleep, patterns and predictability

by

Negin Ashrafi

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science

in the Department of Applied Computer Science

University of Winnipeg

Rough-set based learning methods: A case study to assess the relationship between the clinical delivery of cannabinoid medicine for anxiety, depression, sleep, patterns and predictability

by

Negin Ashrafi

Supervisory Committee

Dr. S. Ramanna, Supervisor
(Department of Applied Computer Science)

Dr. C. J. Henry, Member
(Department of Applied Computer Science)

Dr. A. M. Kumar, External Member
(Department of Information Technology, National Institute of Technlogy Karnataka - Surathkal)

ABSTRACT

COVID-19 is an unprecedented health crisis causing a great deal of stress and mental health challenges in populations in Canada. Recently, research is emerging highlighting the potential of cannabinoids' beneficial effects related to anxiety, mood, and sleep disorders as well as pointing to an increased use of medicinal cannabis since COVID-19 was declared a pandemic. Furthermore, evidence points to a correlation between mental health and sleep patterns. The objective of this research is threefold: i) to assess the relationship of the clinical delivery of cannabinoid medicine, by utilizing machine learning, to anxiety, depression and sleep scores; ii) to discover patterns based on patient features such as specific cannabis recommendations, diagnosis information, decreasing/increasing levels of clinical assessment tools (GAD7, PHQ9 and PSQI) scores over a period of time (including during the COVID timeline); and iii) to predict whether new patients could potentially experience either an increase or decrease in clinical assessment tool scores. The dataset for this thesis was derived from patient visits to Ekosi Health Centres in Manitoba, Canada and Ontario, Canada from January, 2019 to April, 2021. Extensive pre-processing and feature engineering was performed. To determine the outcome of a patients treatment, a class feature (Worse, Better, or No Change) indicative of their progress or lack thereof due to the treatment received was introduced. Three well-known supervised machine learning models (tree-based, rule-based and nearest neighbour) were trained on the patient dataset. In addition, seven rough and rough-fuzzy hybrid methods were also trained on the same dataset. All experiments were conducted using a 10-fold CV method. Sensitivity and specificity measures were higher in all classes with rough and rough-fuzzy hybrid methods. The highest accuracy of 99.15% was obtained using the rule-based rough-set learning method.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

First of all, I would like to thank professor Sheela Ramanna for being so supportive, caring, and understanding throughout my studies at the University of Winnipeg. She has always encouraged and guided me in every step of my academic life. I can't express my gratitude of the opportunity she provided for me to come to Canada and pursue my education. Studying under her supervision and taking advantage of her knowledge and experience is a great privilege for me.

I am thankful of Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants, Mitacs Research Training Award, Mitacs Accelerate Program and Ekosi Health Center which helped me financially.

I would like to thank professor Christopher Henry who is not only my former professor but also one of the committee members and his support was with me since the beginning.

I would like to thank Connie Arnhold, Ferdinand Borillo and other faculty and staff of Graduate Studies and especially members of Applied Computer Science Department for their help and guidance. Special thanks to Karen Debroni and Evan Loster for their guidance in this research.

I would like to thank my friends who supported me through ups and downs of this journey.

Last but not least, I am grateful for my family who has always been there for me and had faith in me. Without their support, I could not have made it this far.

Negin Ashrafi

DEDICATION

To my family for their unconditional support

Chapter 1

Introduction

COVID-19 is an unprecedented health crisis causing a great deal of stress and sleep challenges for populations in Canada. Research is emerging highlighting the potential of cannabinoids' beneficial effects related to chronic pain [2], substance use [3], addiction [4] and poor mental health [5, 6]. Recent studies point to the clinically significant acute impacts the pandemic is having on insomnia rates [7]. Where there is recent research which points to the potential positive impact cannabinoid may have regarding sleep [8, 9], a 2017 review [10] of the literature on cannabinoid, cannabinoids and sleep suggested mixed results and highlighted the need for further research.

With the availability of large amounts of patient data, machine learning (ML) techniques, specifically, supervised and deep learning classifiers, have made it possible to detect, diagnose and treat mental health disorders. Common dataset formats include: Electronic health records (EHR) [11], Social Media (e.g., twitter, Reddit) [12, 13], Image (e.g., MRI) [14] and Audio [15]. [16], present an in-depth review of about 300 papers related to ML and its application in mental health. The most common ML methods used include: support vector machines, decision trees, naive bayes, k-nearest neighbour, and neural networks (deep learning). Latent Dirichlet allocation (LDA) and sentiment analysis methods were used for learning from textual and social media data. Predicting mental health from social media data is an interdisciplinary area also known as *human-centric machine learning* where human insights are combined with data driven predictions [17]. Ethical tensions in inferring mental health states of individuals from social media data are discussed in [18]. In another study [19], Twitter data was used in public health in surveillance, detection, and prevention of events. In [20] XGboost classifier was used

to assess the effectiveness of biomarkers to classify depression cases from healthy cases using a large dataset from Netherlands. In a recent study [21], ML methods were used to assess whether the adolescents with depressive symptoms had access to treatments and, if yes, where the treatments were received. [22] proposed an ensemble of six classifiers to predict general anxiety disorder (GAD) and major depressive disorder (MDD) problems. [23] discuss application of Natural Language Processing techniques to EHR phenotyping (unstructured text) which contain narrative text such as physician notes for improving mental health services. [24] present a survey of papers on mental health detection using ML techniques in Online Social Networks. A review of 2,261 articles on the application of deep learning methods in mental health outcomes was presented in [25]. In this thesis, various deep neural network architectures as well as different forms of clinical data (neuroimages, EMR, audio visual and social media) are discussed.

Since 1991, rough set theory has been applied extensively in medical applications [26, 27]. To the best of our knowledge, there are only a few papers related to the application of rough sets in mental health. In [28], rough set theory was used to explore the relationship between human psychological state (scores of a psychological scale) and physiological state (level of the secretory biomarkers). In [29], the authors use rough sets instead of conventional linear correlation analysis for mining the relationship between a subjective stress scale and salivary cortisol stress biomarker. In [30], a hybrid rough set and Taguchi-genetic algorithm (RS-HTGA) was proposed to determine the relationship between mental stress and biomedical signals. The efficacy of their method was tested on a clinical dataset comprising 362 cases (196 male, 166 female). In [31], the RS-HTGA algorithm achieved sensitivity, specificity, and precision scores of 96%. In [32], the authors present an application of rough sets for attribute reduction to identify depressive episodes.

In this thesis, we seek i) to evaluate the relationship of the clinical delivery of cannabinoid medicine for anxiety, depression and sleep scores by utilizing machine learning; ii) to discover patterns based on patient features such as specific cannabinoid recommendations, diagnosis information, decreasing/increasing levels of clinical assessment tools (GAD7, PHQ9 and PSQI) scores over a period of time (including during the COVID timeline); and iii) to predict whether new patients could potentially experience either an increase or decrease in clinical assessment tool scores. The dataset for this study was derived from patient visits to Ekosi Health

Centres in Manitoba and Ontario, Canada from January, 2019 to April, 2021. Extensive pre-processing and feature engineering was performed on the dataset. To determine the outcome of a patients treatment, a class feature (Worse, Better, or NoChange) indicative of their progress or lack thereof due to the treatment received was introduced. A two-class experiment (Worse or Better) was also explored. Well-known supervised machine learning classification algorithms (tree-based, rule-based, and nearest neighbour) in addition to rough and fuzzy methods were trained on the patient dataset. All experiments were conducted using a 10-fold CV stratified method. Also, prediction of new cases using the rough set-based classifier (LEM2 method) is presented.

The results demonstrate that rough-set based classifier (with LEM2) is superior to all other tested methods in terms of overall classification accuracy, accuracy per class, sensitivity, and specificity values for both the 2-class and the 3-class experiments. A statistical t-test reveals that there is a difference between rough-set based classifier and other tested classifiers for the 3-class experiment. The contribution of this thesis is a novel application of rough and fuzzy classification learning to a case study involving cannabinoid medicine and anxiety, depression, and sleep pattern data [33].

1.1 Thesis Layout

The rest of this thesis organized as follows:

Chapter 2 provides an overview of applications of machine learning techniques in mental health

Chapter 3 provides a theoretical framework for granular computing algorithms

Chapter 4 explains the Ekosi Health Center dataset used as a case study in this thesis and provide details about features

Chapter 5 gives experiments conducted on the medical data set with various algorithms followed by a discussion of the results

Chapter 6 concludes the thesis and provides future research directions

Chapter 2

Related Work-ML in Mental Health

2.1 Overview

In this section, we will review some machine learning classifiers applied on different types of mental health data. In [16], the authors present an in-depth review of 300 papers related to machine learning and its application in mental health. The review includes common mental health conditions such as Alzheimer's disease, depression and schizophrenia. The most common ML methods used were support vector machines, decision trees, naive bayes, k-nearest neighbour and neural networks (deep learning). Latent Dirichlet allocation (LDA) and sentiment analysis methods were used for learning from textual and social media data. Most of the articles focus on detection and diagnosis of mental health in individual patients. There were fewer papers in the domain of public health, treatment and support, as well as research and clinical administration [16].

Social media is a large source of human data that provides information about peoples thoughts, feelings, moods, and experiences. Predicting mental health from social media data is an interdisciplinary area also known as *human-centric machine learning* where human insights are combined with data driven predictions [17]. The authors identify five discourses (patient/disorder, social media, scientific, data/machine learning, person) in 55 papers using a search criteria of 164 terms. The paper discusses the challenges in making a case study as each discourse can happen in other contexts as well. Ethical tensions in inferring mental health states

of individuals from social media data are discussed in [18]. In another study [19], twitter data was used in public health in surveillance, detection and prevention of events.

A review of 2,261 articles on the application of deep learning methods in mental health outcomes is presented in [25]. In this paper, various deep neural network architectures as well as different forms of clinical data (neuroimages, EMR, genetic, audio visual and social media) are discussed. In [34], the authors present a review of 54 papers covering effective systems involving the fusion of human computer interaction (HCI), ML and mental health areas. In [35], the authors present a survey of current state of ML applications in psychotherapy research consisting of 51 studies where text messages between patient and counselor, transcripts of sessions, audio recordings, and patient self-reports were some of the data used for training the ML models. The authors conclude that the size of data is an important factor and suggest that the number of data points should be 10 times more than the number of features.

In section 2.1.1, we discuss a few representative papers that use supervised learning, followed by clustering and semi-supervised methods in section 2.1.2. We conclude the chapter with deep learning methods in section 2.1.3.

2.1.1 Supervised Learning

In [36], Bayesian network was used to select important depression factors. Three Bayesian classifiers were proposed taking into account various factors such as guilt, dangerous behaviour, loss of interest, etc., to calculate the posterior probability and predict the probability of suicide to find the optimum model and significant factors. Prediction of risk of a child developing mental health symptoms as an adolescent is another important issue to screen in order to prevent severe outcomes Certain factors such as environment and personality can affect the classifier performance and logistic regression is the most commonly used classifier to predict mental problems in mid-adolescence [37]. In this study, random forest, XGBoost, neural networks and SVM classifiers were used. The results demonstrate a very slight improvement with random forest and SVM classifiers compared to the logistic regression classifier using the area under the receiver operating characteristic (AUC) measure. In a recent study [21], ML methods were used to assess whether adolescents with

depressive symptoms had access to treatments and if yes, where the treatments were received. Random forest regressor and linear regression (elastic nets) were used to perform sensitivity analyses. Elastic nets outperformed random forests and out of 1,671 adolescents, 53.38% had access to treatment during a 12-month period.

In [20] XGboost classifier was used to assess the effectiveness of biomarkers to classify depression cases from healthy cases using a large dataset from Netherlands. The paper attempts to find an association between important biomarkers (features) and depression in order to address the problem of patients not disclosing depressive symptoms during a clinical interview.

In [11], features such as blood report, physical demographic, sleep history extracted from electronic health records were used to detect obstructive sleep apnea disorder using an SVM classifier with 68.06% score for accuracy measure and 88.76% score for sensitivity measure.

Authors of [22] proposed a novel machine learning pipeline containing 6 Classifiers (XGBoost, Random Forest, Support Vector Machine, KNN and a neural network tuned using Bayesian hyperparameter optimization) to predict general anxiety disorder (GAD) and major depressive disorder (MDD) problems. 59 features from electronic health records of students were used. The most predictive features for MDD were, satisfaction of living conditions and having public health insurance. The top effective features for GAD were, vaccinations being up to date and marijuana use.

2.1.2 Other methods

In [38] three clustering methods (k-means, agglomerative hierarchical and K-mediods) were first used to label the target population as mentally distressed, neutral, and happy. Clustering results were validated using mean opinion score. Seven well-known classifiers were used with SVM, k-nearest neighbour and random forest classifiers giving the best result (90% accuracy).

In [12], a semi-supervised method was used to classify patients into 4 categories (anxiety, depression, bipolar, and ADHD) based on social media posts on the Reddit website. Co-training method using Random Forest, Support Vector Machines and Naive Bayes classifiers resulted in better performance. First, a small labelled dataset was used to train weak classifiers. Next, these classifiers were used in an

iterative manner to generate labelled examples from a larger unlabelled dataset.

2.1.3 Deep Learning

In [39], multilayer feedforward deep neural network was applied to unstructured text notes in EMRs consisting of 861 documents related to 366 patients over a period of six months. The study in youth depression was meant to phenotype potential participants for research recruitment. Brute Force method was used to to label the documents as positive (inclusion criteria) or negative (exclusion criteria). 3 deep models were trained on the labelled documents using TF-IDF vectorization.
In [40], Long Short-Term Memory (LSTM) was used to predict the risk of future depressive episodes and intervention recommendation from illness trajectories (including time-series data) extracted from EMRs. Mental health and diabetes cohorts were used as case studies. For baseline comparison, SVM and Random Forests were used on non-temporal features.
In [14], the authors present a survey of popular deep learning methods such as convolutional neural networks, belief networks, recurrent networks, probabilistic networks to detect neurological disorders (Parkinson's, Alzheimer's and schizophrenia) using magnetic resonance imaging (MRI) datasets. The review concludes that the CNN model outperforms other deep learning methods.
In [41], Three-dimensional convolutional neural networks (3D-CNNs) were applied using magnetic resonance imaging data for diagnosing patients with Alzheimers disease. Three different CNN models (both 2D and 3D) were trained to generate feature maps. Two different classifiers were incorporated (softmax and SVM) with 3D-CNN-SVM giving the best result.

2.1.4 Clinical Datasets: Types and summary list of related papers

Fig. 2.1 shows the different forms of clinical data sets used in many of the surveys. The following tables give a summary list of reviewed papers grouped by the type of clinical dataset used in various mental health studies.

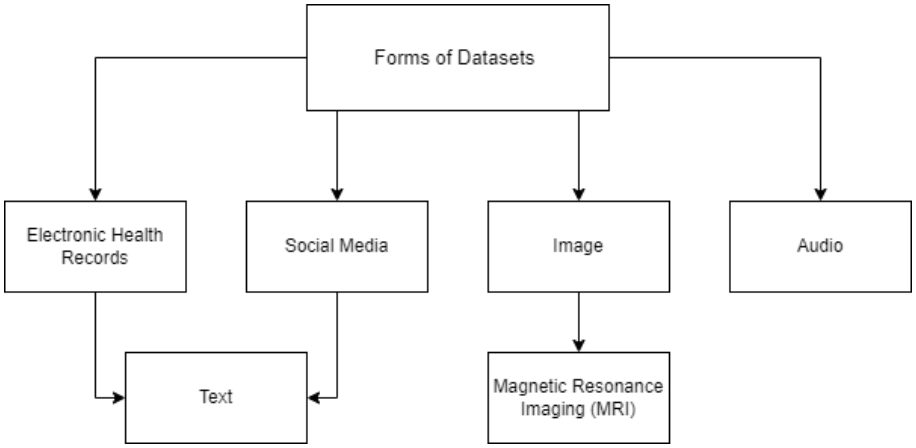


Figure 2.1: Clinical data sets

Table 2.1: EHR as dataset - related papers

Improving diagnosis of depression with XGBOOST machine learning model and a large biomarkers Dutch dataset [20]
Towards Validating the Effectiveness of Obstructive Sleep Apnea Classification from Electronic Health Records Using Machine Learning [11]
Predicting mental health treatment access among adolescents with elevated depressive symptoms: Machine learning approaches [22]
Predicting healthcare trajectories from medical records: A deep learning approach [40]
Applying deep neural networks to unstructured text notes in electronic medical records for phenotyping youth depression [39]
The use of electronic health records for psychiatric phenotyping and genomics [42]
A machine learning approach to modeling PTSD and difficulties in emotion regulation [43]
Predicting personalized process-outcome associations in psychotherapy using machine learning approachesA demonstration [44]

Table 2.2: Social media as dataset - related papers

A taxonomy of ethical tensions in inferring mental health states from social media [18]
Who is the" human" in human-centered machine learning: The case of predicting mental health from social media [17]
A scoping review of the use of Twitter for public health research [19]
A novel co-training-based approach for the classification of mental illnesses using social media posts [12]
Predicting social anxiety treatment outcome based on therapeutic email conversations [45]
Large-scale analysis of counseling conversations: An application of natural language processing to mental health [46]

Table 2.3: Images as dataset - related papers

Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimers disease, Parkinsons disease and schizophrenia [14]
Application of machine learning to structural connectome to predict symptom reduction in depressed adolescents with cognitive behavioral therapy [47]
Using fMRI and machine learning to predict symptom improvement following cognitive behavioural therapy for psychosis [48]
Individualized treatment response prediction of dialectical behavior therapy for borderline personality disorder using multimodal magnetic resonance imaging [49]
Deep learning based diagnosis of Parkinsons disease using convolutional neural network [50]

Table 2.4: Audio as dataset - related papers

A technology prototype system for rating therapist empathy from audio recordings in addiction counseling [15]

Chapter 3

Data Preparation

3.1 Overview

The original dataset includes 541 unique patients and 32,514 records (for single and multiple visits). In this thesis, patients with at least two different dates of a medical appointment with one of the Health Centres were considered (referred to as multiple visit dataset). The ages for youngest and oldest patients were 6 and 108 years respectively (with a mean value of 58.61). Additionally, this multiple dataset included 390 types of diagnoses with 75 unique cannabidiol formulations. After data cleaning, diagnoses types that were not of interest in this study removed, the multiple visit dataset was reduced to 354 patients from 375 patients.

- Patient Id : Since this feature uniquely identifies a patient, due to privacy reasons, this feature value was anonymised by removing each patient's name, date of birth and any information that might reveal the patient's identity.
- Age: This feature gives the age of the patient where the minimum value for age is 6 and the maximum value is 108.
- Clinical Assessment Tool (CAT): This feature indicates the type of the clinical measure assessment tool that was utilized to assess and score the patient. Three specific CAT types were observed in this study; the GAD-7 (General Anxiety Disorder 7), PHQ-9 (Patient Health Questionnaire 9), and PSQI (Pittsburgh Sleep Quality Index).
- CAT Value : The feature gives the values for each of the CAT types: GAD-7, PHQ-9, and PSQI.

- CAT Observation Date: This feature gives the date on which a CAT value was observed.
- Sex Id: This feature gives the gender and the distribution of the patients coded as 1: male (34.2%) and 2: female (65.8%).
- Cannabinoid recommendation: This feature indicates the specific cannabinoid recommendation. The medical cannabis products contain varying amounts of cannabidiol (CBD) and tetrahydrocannabinol (THC), two phytocannabinoids found in cannabinoid.
- Diagnosis: This feature indicates the diagnosis of the patient. There were 390 types of diagnoses and only 13 types were considered in this thesis.

The raw data had several problems such as missing or invalid values, continuous values for dosage and similar diagnosis which required extensive preprocessing. In the following section, we discuss the preprocessing steps applied to the dataset.

3.1.1 Preprocessing

The final dataset after preprocessing for experimentation was: 8,281 records (2,911 male and 5,730 female). The description of the steps are as follows:

- Invalid and missing values: Invalid and null values were found in gender and CAT value features and were removed. For example, there were 114 records that gender had a value other than 1 or 2. Also, in the original dataset, there were 18 records that CAT value greater than 27. There were very few records with missing values which were also removed.
- Diagnosis coding: The raw data consisted of 390 diagnoses categories. Some low occurring or categories not relevant to this study were removed (E.g.: ADHD, MS, Anemia, Vitiligo, Blood Clot, Schizophrenia, and Overweight). Other granular categories such as migraine, classical migraine, common migraine, chronic migraine without aura were combined into the broader migraine category. In this study, we were primarily interested in chronic pain, so patients with migraine and headache were included in the chronic pain category.

- Cannabinoid recommendation coding: The values for this feature were continuous since they represent dosage values. Since we were only interested in a broad class of values, these values were converted into integers using regular expressions (using Python regular expression package).

- Multiple cannabinoid recommendations: Many patients (almost 40%) were recommended more than one cannabinoid product for one particular diagnosis in a single visit. This was primarily for cannabinoid product classes CBD and CBD AND THC:CBD. For such patients, the recommendation was changed to CBD AND THC:CBD (category 3). This resulted in duplicate records and these duplicate records were removed.

- Multiple CAT values: Some patients had a different value for GAD-7/PHQ-9/PSQI during a single visit. For this feature, records with largest CAT value (most severe) were recorded.

- Time of visit: All time values with a small difference during a single visit were standardized and 21 patients had a slight time difference in at least one record.

- CAT value coding: This generated feature was designed to merge CAT value and CAT types: A0-A3, D0-D4, and S0-S3 to represent anxiety (GAD-7), depression (PHQ-9), and sleep disorder (PSQI) severity level respectively.

Table 3.1 provides scores/severity levels for each of the Clinical Assessment Tools and their diagnostic status that was used for our experiments.

Table 3.1: Scores/severity levels for each CAT Value and their diagnostic status

CAT Type	CAT Value Range(score)	Diagnostic Status
GAD-7	0 - 4	Minimal Anxiety
	5 - 9	Mild Anxiety
	10 - 14	Moderate Anxiety
	15 - 21	Severe Anxiety
PHQ-9	0 - 4	Minimal Depression
	5 - 9	Mild Depression
	10 - 14	Moderate Depression
	15 - 19	Moderately Severe Depression
	20 - 27	Severe Depression
PSQI	0 - 4	Good Sleep quality
	5 - 9	Poor Sleep Quality
	10 - 15	Bad Sleep Quality
	16 - 20	Severely Bad Sleep Quality
	21	Terrible Sleep Quality

One of the main objectives of this study was to detect patterns in the fluctuations of values for GAD-7, PHQ-9 and PSQI (clinical assessment tools) for a patient during a time period. Figure. 3.1 shows the number of patients from 2019 to 2021. As the figure shows, the number of patients that visited the Centre was the highest (59) during April 2020 which was also the start of the first wave of COVID. In particular, we were interested in the overall outcome of a patients quality of life in terms of whether their GAD-7/PHQ-9/PSQI scores were increasing/decreasing/constant during the period of observation. In addition, this information had to be co-related with their cannabinoid product recommendation and diagnosis.

Figure. 3.2 shows the trends in score values for a single patient at the peak of COVID. It can be seen that in Figures. 3.2a and 3.2b there is no regular pattern for GAD-7/PHQ-9/PSQI scores.

3.1.2 Engineered Feature - Patient Status

To determine the outcome of a patients treatment, we introduced a new feature (status) indicative of their progress or lack thereof due to the treatment received over a period of time. Three values for status were decided: Worse, Better, or NoChange. An additional reason for introducing these labels was to train

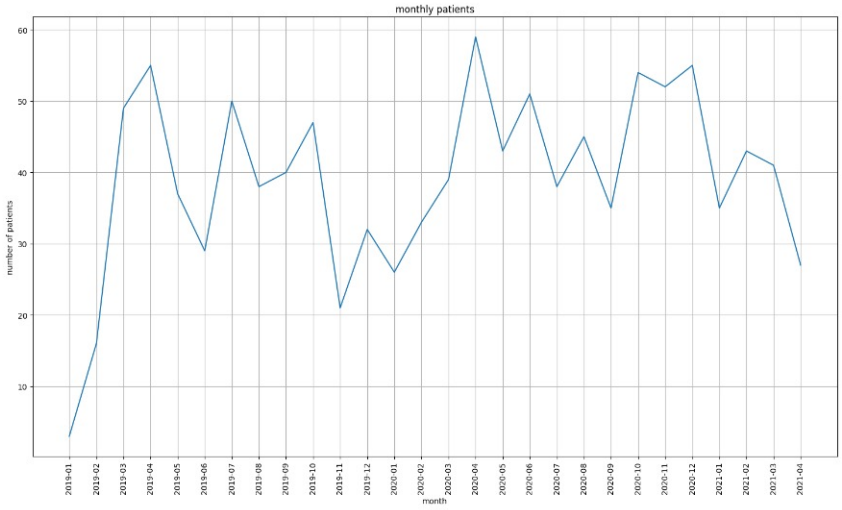
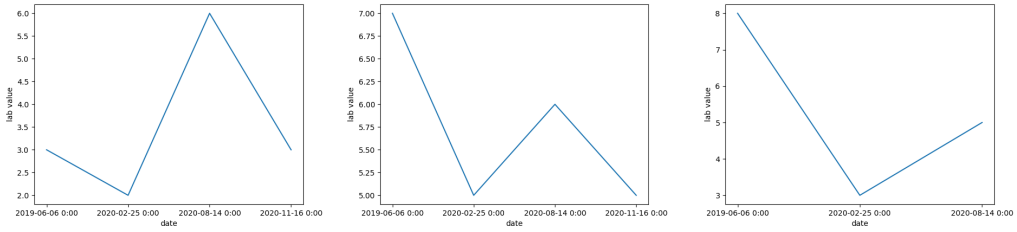


Figure 3.1: The number of patients distribution in each month



(a) GAD-7 (b) PHQ-9 (c) PSQI

Figure 3.2: The observed scores (CAT value) for one patient

classification models so that these models can be used to determine (or predict) the status of a new patient. Figure 3.3 provides the flowchart for computing the value of this feature.

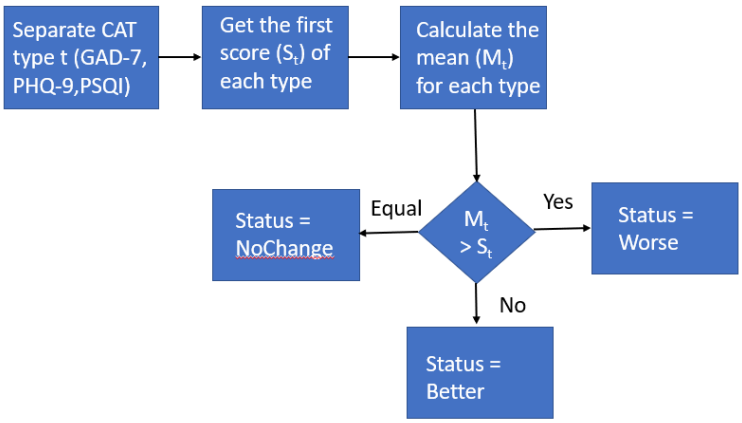
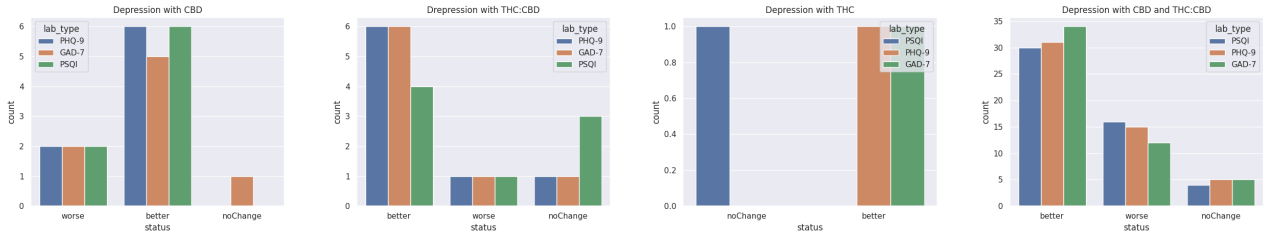


Figure 3.3: FlowChart - Patient Status Computation

The assumption behind this computation was that, since the score values for a disorder type does not follow any trend (as shown in Fig. 3.2), a mean score value would be representative of a patients score over the entire time period. In addition, there were unequal scores recorded for each patient during a time period. This problem was also observed for different disorder types as well. Hence, we separated the data into different CAT types first and then performed the labelling. This method also solved the problem of lack of observations of a CAT type with a time period for any given patient.

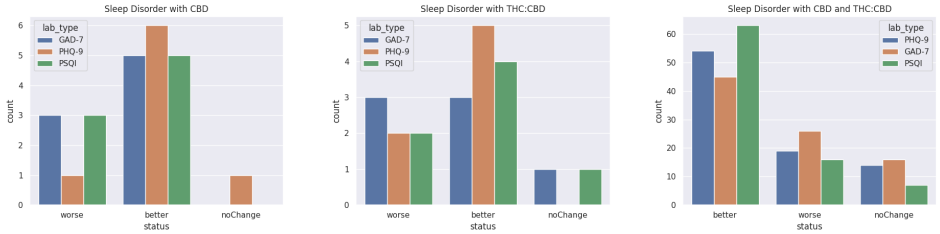


(a) CBD (b) THC:CBD (c) THC (d) CBD and THC:CBD

Figure 3.4: Distribution of patients with their status, diagnosed with Depression

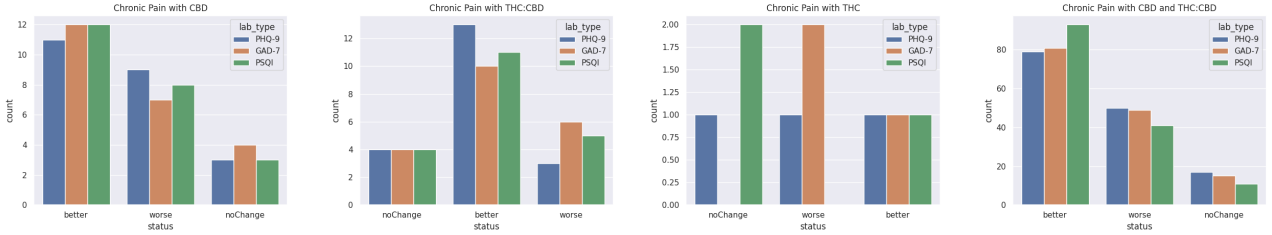
Figure 3.4, shows the distribution of patient records based on i) labelled patient's status (Worse, Better, NoChange) , ii) diagnosis (depression), and iii) CAT type for the four different types of cannabinoid formulations: CBD, THC:CBD, THC and

CBD AND THC:CBD. The distribution of patients for other diagnoses (ex: Sleep Disorder, Chronic Pain, Arthritis, Anxiety) can be found in Figure 3.5, 3.6, 3.7, 3.8. However, chronic pain is the most frequent diagnosis and there were no patients with sleep disorder diagnosis who were recommended THC formulation. The distribution of diagnosis and cannabinoids product recommendation classes at the beginning (before changing cannabinoids product recommendation and removing duplicates) are given in Figures 3.9, 3.10 respectively.



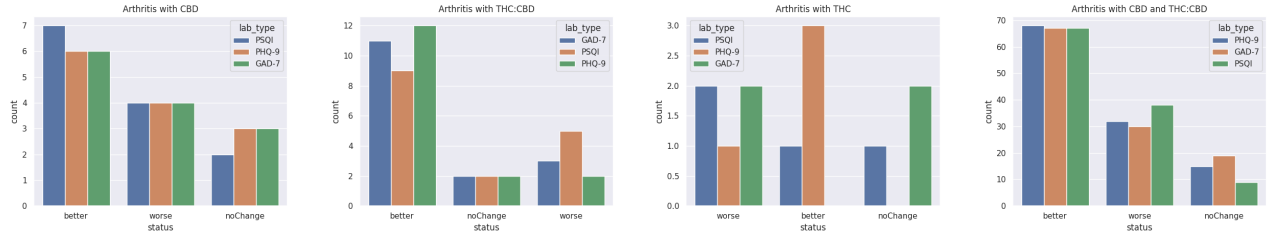
(a) CBD (b) THC:CBD (c) CBD and THC:CBD

Figure 3.5: Distribution of patients with their status, diagnosed with Sleep Disorder



(a) CBD (b) THC:CBD (c) THC (d) CBD and THC:CBD

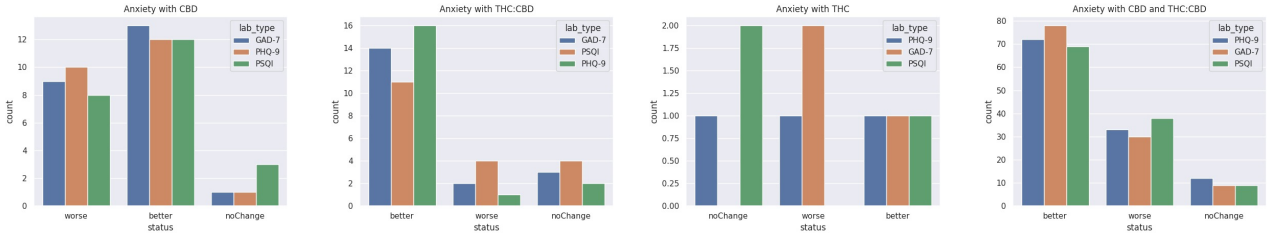
Figure 3.6: Distribution of patients with their status, diagnosed with Chronic Pain



(a) CBD (b) THC:CBD (c) THC (d) CBD and THC:CBD

Figure 3.7: Distribution of patients with their status, diagnosed with Arthritis

Figures 3.5, 3.6, 3.7 and 3.8 show the distribution of patient records based on i) labelled patient's status (worse, better, noChange) , ii) diagnosis, and iii) CAT type for the four different types of cannabinoids formulations: CBD, THC:CBD, THC and CBD AND THC:CBD . Also, note in Fig. 3.5, there were no patients with sleep disorder diagnosis who were recommended THC formulation.



(a) CBD (b) THC:CBD (c) THC (d) CBD AND THC:CBD

Figure 3.8: Distribution of patients with their status, diagnosed with Anxiety

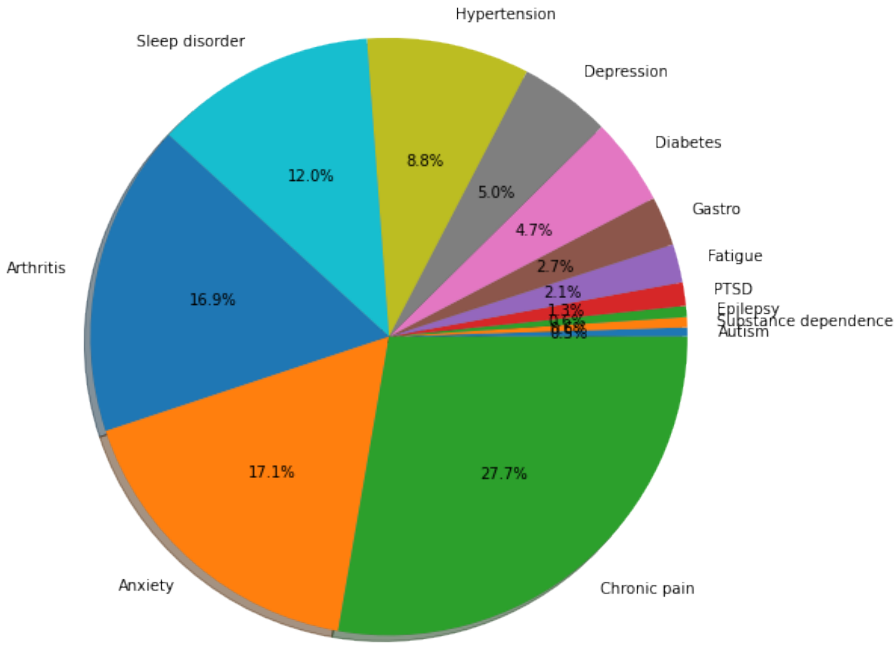


Figure 3.9: The distribution of patient diagnosis classes before data cleaning

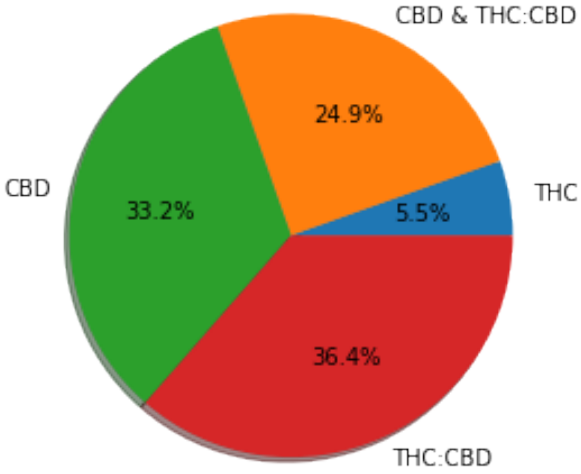


Figure 3.10: The distribution of patient cannabinoids product recommendation classes before data cleaning

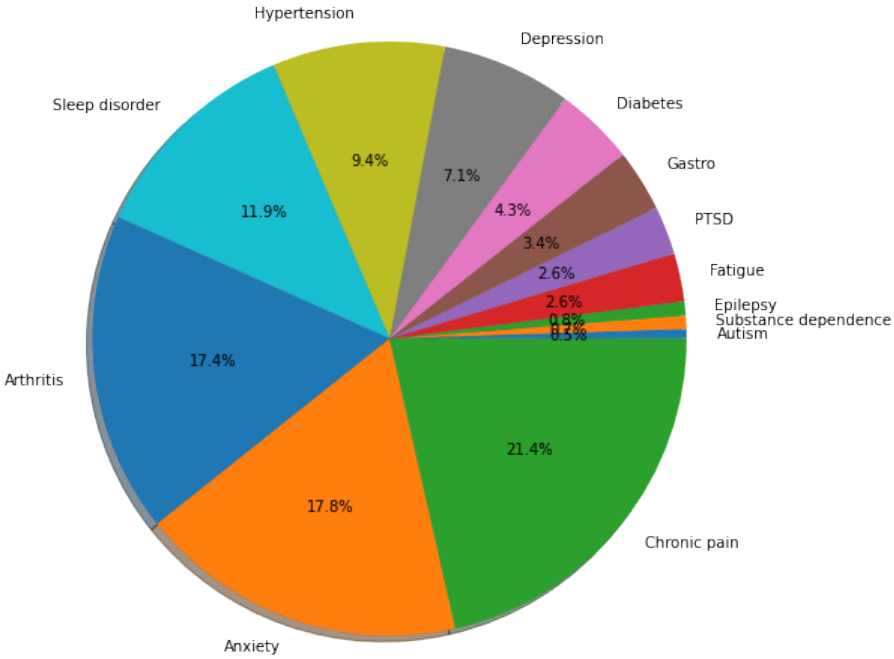


Figure 3.11: The distribution of patient diagnosis classes after data cleaning

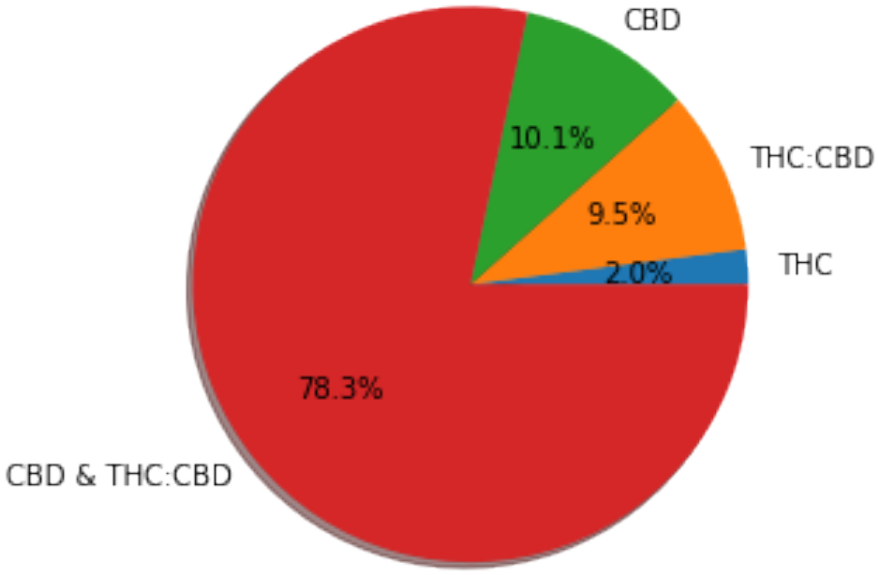


Figure 3.12: The distribution of patient cannabinoids product recommendation classes after data cleaning

Chapter 4

Formal Models: Rough Sets and Fuzzy Rough Sets

4.1 Preliminaries

In this section, we present a brief review of rough and fuzzy rough set theory concepts that were used in this thesis. Specifically, we use different forms of fuzzy and rough nearest neighbor classification algorithms.

4.1.1 Rough Sets

In classical set theory, we can classify whether elements either belong to a set or not. This is a precise or crisp set where the sets have sharp boundaries. However, when boundaries are *unsharp or vague*, it is difficult to classify elements uniquely to one set. In other words, this will result in a boundary region with elements that cannot be classified precisely. Rough set theory was proposed by Zdzislaw Pawlak in early 80's as a mathematical framework to analyze vague data and ill-defined objects based on an indiscernibility or equivalence relation [51, 52]. Equivalence relations generate equivalence classes and the notion of indiscernibility is defined relative to a given set of attributes [27]. Due to the lack of knowledge (or uncertainty) that objects might belong to more than one set (or class), two approximation operators (*lower* and *upper*) are introduced in rough set theory to generate precise sets. In supervised classification, the advantage of rough set theory is that no prior or additional data is needed to categorize data into classes [26].

Figure 4.1 shows the regions that emerge with rough set approximation. The lower approximation consists of the objects that certainly belong to the set (orange region) and upper approximation consists of objects that their membership is not certain (green region). The regions are depicted as squares only for the sake of illustration, but they can be of arbitrary shape. We should note that each granule can contain an arbitrary number of objects or may be empty. The oval denotes the target X which, in the case of supervised learning, is either a class or a pattern that needs to be learned.

Let U be a finite, non-empty universe of objects and let $R \subseteq U \times U$ denote a binary relation on the universe U. R is called an *indiscernibility relation* and for rough sets, it has to be an *equivalence relation*. The pair $(U, R) = \mathcal{A}$ is an *approximation space* $\mathcal{A}$ [53]. Let $X \subseteq U$ be a target concept in this universe. Then the task is to create an approximated representation for X in U with the help of R. Let $[x]_R$ denote the indiscernibility class of x i.e. $y \in [x]_R \iff (x, y) \in R$. Then, every equivalence class forms a *granule* or *partition* containing objects that are indiscernible for this approximation space $\mathcal{A}$. Therefore, every single item in a granule is considered identical and inseparable. These granules are approximated by the following means:

- **Lower approximation**. Intuitively, these are the objects which *certainly* belong to X with respect to $\mathcal{A}$.

$$\mathcal{L}_{\mathcal{A}}(X) = \{x \in U : [x]_R \subseteq X\}.$$

- **Upper approximation**. Intuitively, these are the objects which *may* belong to X with respect to $\mathcal{A}$.

$$\mathcal{U}_{\mathcal{A}}(X) = \{x \in U : [x]_R \cap X \neq \emptyset\}.$$

These two approximations will also form the following two regions:

- **Boundary region**. These are the objects occurring in the upper approximation but not in lower approximation of X.

$$\mathcal{B}_{\mathcal{A}}(X) = \mathcal{U}_{\mathcal{A}}(X) - \mathcal{L}_{\mathcal{A}}(X).$$

- **Negative region**. These are the objects that certainly don't belong to X.

$$U - \mathcal{U}_{\mathcal{A}}(X).$$

With this framework, we have two different types of sets: a set X is called a *crisp set* if and only if $\mathcal{B}_{\mathcal{A}}(X) = \emptyset$. Otherwise, it is called a *rough set*. The pair $(\mathcal{U}_{\mathcal{A}}(X), \mathcal{L}_{\mathcal{A}}(X))$ forms the *rough approximation* for X (see Figure 4.1 for an illustration).

4.2 Rough Sets: Upper and lower approximations

Figure 4.1 shows the regions that emerge with rough set approximation.

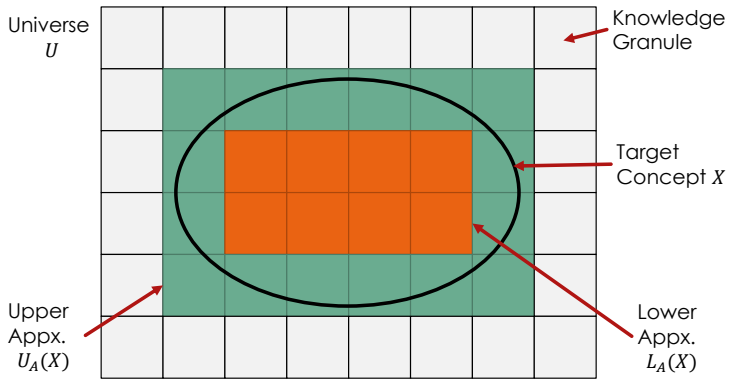


Figure 4.1: Rough Sets: Upper and lower approximations [1]

Table 4.1 gives a sample patient data with three features (CAT type and value are combined) and one decision feature for illustration.

We illustrate the lower and upper approximations in terms of sample patient data presented in Table 4.1.

Example 1. *Let* $X = \{x1, x3, x6, x7, x8, x10\}$ *for decision Better.*
Let $[x]_R = \{x1, x2, x7, x3, x4, x6, x5, x8, x9, x10\}$ *and*
$A = \{$*CAT type-value, Diagnosis, Cannabinoids recommendation*$\}$

$$\mathcal{L}_{\mathcal{A}}(X) = \{x1, x3, x8, x10\}.$$

Table 4.1: Sample Patient Data Table

Sample	CAT type/value	Diagnosis	Cannabinoids recommendation	Status (decision)
x1	D2 (PHQ-9, 10-14)	Hypertension	CBD & THC: CBD	Better
x2	A3 (GAD-7, 15-21)	Chronic Pain	CBD	Worse
x3	S2 (PSQI, 10-15)	Fatigue	CBD & THC: CBD	Better
x4	D3 (PHQ-9, 15-19)	Sleep Disorder	CBD & THC: CBD	Worse
x5	A0 (GAD-7, 0-4)	Sleep Disorder	CBD & THC: CBD	Worse
x6	D3 (PHQ-9, 15-19)	Sleep Disorder	CBD & THC: CBD	Better
x7	A3 (GAD-7, 15-21)	Chronic Pain	CBD	Better
x8	D2 (PHQ-9, 10-14)	Anxiety	THC: CBD	Better
x9	A3 (GAD-7, 15-21)	Depression	CBD & THC: CBD	Worse
x10	D0 (PHQ-9, 0-4)	Hypertension	CBD & THC: CBD	Better

$$\mathcal{U}_{\mathcal{A}}(X) = \{x1, x3, x4, x6, x2, x7, x8, x10\}.$$

$$\mathcal{B}_{\mathcal{A}}(X) = \{x4, x6, x2, x7\}.$$

4.2.1 Fuzzy Rough Sets

Fuzzy set theory was proposed by Lotfi Zadeh [54] as an extension of traditional set theory to deal with uncertainty and vagueness. In the context of fuzzy sets, let X denote the universe, a fuzzy set $A \in X$ is characterized by a mapping X $\rightarrow [0, 1]$ which is also called a membership function. A fuzzy relation R in X which is also a fuzzy set and is characterized by a mapping R: X $\times$ X $\rightarrow [0, 1]$ [54]. [55, 56, 57], introduced the idea of combining fuzzy and rough sets to develop soft similarity classes i.e., fuzzifying the approximations of rough set theory. Formally, a fuzzy rough set is a pair $(A_1, A_2) \in (X, R)$ where A is a fuzzy set in X such that $R \downarrow A = A_1$ and $R \uparrow A = A_2$ and R is a fuzzy relation in X [58]. Fuzzy rough sets permit partial membership of an object to the lower and upper approximations and the approximate nature of information are modeled by means of fuzzy indiscernibility relations. In general, R can be considered as a fuzzy tolerance relation such $R(x, x) = 1$ and $R(x, y) = R(y, x)$ for all x, y in X. Let U be the universe and R the fuzzy tolerance relation in U which is a mapping $U \rightarrow [0, 1]$ and A is a fuzzy set in U, the upper $(R \uparrow A)$ and lower approximation of A $(R \downarrow A)$ is calculated by R using different methods. The general form for this calculation from [59] is as follows:

$$(R \downarrow A)(x) = \inf_{y \in u} \mathcal{I}(R(x, y), A(y)) \tag{4.1}$$

$$(R \uparrow A)(x) = \sup_{y \in u} \mathcal{T}(R(x, y), A(y)) \tag{4.2}$$

where $\mathcal{I}$ is an implicator and $\mathcal{T}$ is a t-norm which are fuzzy logic connectives crucial for fuzzy rough hybridization. The Kleen-Diennes Implicator implemented in the WEKA platform is defined as

$$\mathcal{T}_M = min(x, y) \tag{4.3}$$

$$\mathcal{I}_M = max(1 - x, y) \tag{4.4}$$

The FRNN algorithm combines fuzzy rough approximations and fuzzy nearest neighbour algorithm which forms the basis for classification learning. Fuzzy nearest neighbour (FNN) algorithm was introduced to classify test examples based on their similarity to a given number K neighbours of training examples and membership degrees to (crisp or fuzzy) class labels of these neighbours [60]. In the Fuzzy Rough Nearest Neighbour (FRNN) implementation in WEKA, given a set of conditional attributes C, R is defined as where R_a is the degree to which objects x and y are similar for attribute a [59]:

$$R(x, y) = \min_{a \in C} R_a(x, y) \tag{4.5}$$

The two options for R_a referred to option 1 (Eqn 4.6) and option 2 (Eqn 4.7) respectively are:

$$R_a^1(x, y) = \exp\Big(- \frac{(a(x) - a(y))^2}{2\sigma_a^2} \Big) \tag{4.6}$$

$$R_a^2(x, y) = 1 - \frac{\|a(x) - a(y)\|}{|a_{max} - a_{min}|} \tag{4.7}$$

where σ_a^2 is variance of attribute a, and a_{max} and a_{min} are maximal and minimal values of attribute a. For the sake of completeness, we use the FRNN algorithm presented in [59] which is implemented in WEKA. We have used option 2 to report the results in Chapter 5.

4.2.2 LEM2

The algorithm LEM2, is a data mining component of LERS system (Learning from Examples using Rough Sets) which uses the idea of blocks of attribute-value pairs. The input to the LEM2 Algorithm is a lower or upper approximation of a target concept [61]. LEM2 generates rules by computing a single local covering of each concept from the decision table [62]. For an attribute-value pair $(a, v) = t$, a block of t, denoted by $[t]$, is a set of all cases from U for attribute a having value v. Let B

Algorithm 1: Fuzzy Rough Nearest Neighbor Classification [59]

Input : A decision table $\mathcal{A} = (U, A \cup d)$ as training data; y is new data
Output: :Class is a predicted class

```
N←NN (y, K); /*NN is the k-nearest neighbor algorithm */
T ←0; Class← ∅; /*T from Eqn. 4.2*/
foreach C ∈ d do
    if ((R ↓ C)(y) + (R ↑ C)(y))/2 ≥ T then
        Class ← C;
        T ← ((R ↓ C)(y) + (R ↑ C)(y))/2;
        end
    Output: Class
```

be a non-empty lower or upper approximation of a target concept represented by a decision-value pair (d, w) where d is the decision with values w [61]. Set B depends on a set T of attribute-value pairs $t = (a, v)$ if and only if $t \in T$, $\emptyset \neq [T] = \cap[t] \subseteq B$

Set T is a minimal complex of B if the following conditions hold [63]:

- $T \neq \emptyset$
- $[T]_U \neq \emptyset$
- $[T]_U \subseteq X$
- there exists no T such that $T \subseteq T$ and T is a minimal complex.

4.2.3 JRIP

JRIP performs Repeated Incremental Pruning to Produce Error Reduction (RIPPER). JRip is a bottomup method rule learner [64]. The algorithm has 4 stages: Growing a rule, Pruning, Optimization and Selection [65].

The algorithm starts from the less prevalent class to the most frequent class and each time grows one rule by adding antecedents (or conditions) to the rule until the rule has maximum accuracy and meets the minimum rule description threshold. The algorithm tries every possible value of each attribute and selects the condition with highest information gain. Then, the algorithm prunes each rule incrementally. After generating the initial rule set, we can generate and prune two variants of each rule. In the next step, the smallest possible description length for each variant and

the original rule is computed. The variant with the minimal description length is selected as the final representative in the rule set.

Chapter 5

Experiments, Results and Discussion

5.1 Overview

In this chapter, we discuss the implementation of machine learning algorithms on the multiple visits data set and analyze the results.

5.2 Experiments

Results of the following algorithms are reported in Table 5.1: Random Forest, JRIP-Ripper Algorithm in WEKA 3.7.2 [1] as well the classical rough sets model Rough Sets implemented in RSES 2.2.2 [66]. The LEM2 algorithm was used to generate the classification results [67]. Results from other fuzzy and rough sets algorithms implemented in WEKA are reported in Table 5.5, we provide the results (average values) in terms of classification accuracy (%), sensitivity(%), specificity (%), and area under the curve (AUC) (%) for the four classifiers.

For the Random Forest classifier, the following parameters were used: maximum depth was set to 6 and number of trees was set to 10. For the JRIP classifier, one fold was used for pruning and two folds for growing the rules. For the FRNN classifier, 10 nearest neighbours were chosen, with Kleen-Diennes Implicator and Kleen-Diennes t-norm. For the results reported in Table 5.1, ten sets of training and testing pairs were used for experimentation across both platforms (10-Fold Cross

[1] https://www.cs.waikato.ac.nz/ml/weka/index.html

Table 5.1: Results - binary and ternary class experiments.

Metric	Fuzzy Rough NN	Random Forest	JRIP	Rough Se
Mean Accuracy Overall (Binary)	97.11	96.22	97.16	99.20
Mean Accuracy Overall (Ternary)	96.79	95.82	96.52	99.34
Accuracy (Binary- Better)	97.9	97.6	99.1	99.4
Accuracy (Binary- Worse)	95.7	93.9	93.8	98.78
Accuracy (Ternary- Better)	97.0	96.3	96.6	99.3
Accuracy (Ternary- Worse)	97.3	96.7	97.6	99.4
Accuracy (Ternary- NoChange)	99.2	98.5	98.7	99.8
Sensitivity (Binary- Better)	97.9	97.6	99.1	99.4
Sensitivity (Binary- Worse)	95.7	93.9	93.8	98.78
Specificity (Binary- Better)	95.7	93.9	93.8	98.78
Specificity (Binary- Worse)	97.9	97.6	99.1	99.4
AUC (Binary- Better)	0.99(96.845)	0.98(95.775)	0.96(96.505)	0.99
AUC (Binary- Worse)	0.99(96.845)	0.98(95.775)	0.96(96.505)	0.99
Sensitivity (Ternary- Better)	97.9	98.1	99.1	99.6
Sensitivity (Ternary- Worse)	95.0	93.6	93.5	98.7
Sensitivity (Ternary- NoChange)	94.2	87.7	86.5	99.3
Specificity (Ternary- Better)	95.5	93.6	92.4	98.9
Specificity (Ternary- Worse)	98.2	98.0	99.4	99.7
Specificity (Ternary- NoChange)	99.6	99.5	99.7	99.9
AUC (Ternary- Better)	0.91(96.7)	0.98(95.85)	0.96(95.75)	99.25
AUC (Ternary- Worse)	0.9(96.6)	0.98(95.8)	0.97(96.45)	99.2
AUC (Ternary- NoChange)	0.95(96.9)	0.96(93.6)	0.95(93.1)	99.6

Validation (CV) stratified method). We considered two forms of outcome of a patients treatment: 2-class (Better or Worse) and 3-class (Worse, Better, or NoChange) referred to as binary and ternary respectively. For the 2-class experiment, the Better class contains 5,157 records and the Worse class contains 3,124 records. For the 3-class, the Better class contains 5,157 records, Worse class contains 2,470 records and the NoChange class contains 654 records. The average number of rules for JRIP for binary classification was 157 and for ternary classification was 208. For RSES, the average number of rules for JRIP for binary classification was 2,843 and for ternary classification was 2,758. The execution time in secs per fold for each classifier was: Fuzzy Rough NN (0.01), Random Forest (0.19-0.25), JRIP (1-1.5) and Rough Sets (12).

The result can be slightly different when we give the whole dataset to the algorithm

Table 5.2: Experiments with automatic 10-fold cross validation

Classes	FRNN(option2)	FRNN(option1)	Random Forest	JRIP	Rough Set
Binary	98.184	96.159	94.36	95.64	98.4
Ternary	96.437	95.894	94.80	96.473	97.5

Table 5.3: Experiments with 10-fold cross validation for 2-class dataset

Fold number	Fuzzy Rough NN	Random Forest	JRIP (WEKA)	Rough Set (RSES)
1	97.82	95.77	95.65	99.3
2	97.7	96.61	98.3	99.3
3	96.85	96.25	97.34	99.6
4	97.1	96.25	98.18	99.1
5	97.34	96.61	97.34	99.3
6	97.1	96.01	97.22	98.8
7	96.61	96.37	97.34	99.0
8	95.89	96.01	97.58	99.3
9	97.58	95.89	96.13	99.6
10	97.1	96.37	96.49	98.7
Mean	97.11	96.22	97.16	99.2

and choose cross validation *default* option since the folds will be chosen differently and randomly. Table 5.2 shows the accuracy for 2-class and 3-class datasets in this case.

Table 5.5 presents classification results of various nearest neighbour implementations using different forms of fuzzy and rough sets.

All results are based on 10-fold CV done automatically by the WEKA FRNN tool. Here is a brief description of the methods.

- Vaguely Quantified Nearest Neighbour (VQNN) classifier [68] which is based on Vaguely Quantified rough sets (VQRS) algorithm introduced in [69]. VQRS uses fuzzy linguistic quantifiers such as *most* and *some* to decide to what extent an object belongs to the lower and upper approximations.
- Fuzzy Unordered Rule Induction Algorithm (FURIA) FURIA extends the well-known RIPPER [70] algorithm and learns fuzzy rules.

Table 5.4: Experiments with 10-fold cross validation for 3-class dataset

Fold number	Fuzzy Rough NN	Random Forest	JRIP (WEKA)	Rough Set (RSES)
1	97.22	95.29	97.34	99.1
2	97.34	96.01	95.04	99.4
3	96.73	96.13	95.53	99.6
4	96.73	95.16	96.98	99.3
5	97.34	96.61	96.61	99.6
6	96.37	95.41	97.46	99.6
7	96.98	95.77	97.46	99.0
8	96.13	96.13	96.13	99.4
9	96.73	95.65	96.61	99.3
10	96.37	96.01	96.014	99.1
Mean	96.79	95.82	96.52	99.34

Table 5.5: Results - WEKA FRNN implementations

Algorithm	Accuracy (3-class)	Accuracy (2-class)	Time 3 class (secs)	Time 2 class (secs)
FNN	78.13	81.17	0.01	0.03
FNN-O	89.51	90.69	0.01	0.01
VQNN	81.99	84.70	0.01	0.02
FURIA	95.17	96.60	60.8	61.4
DC	91.57	92.21	0.01	0.01

- Fuzzy nearest neighbour (FNN) algorithm was introduced to classify test examples based on their similarity to a given number K neighbours of training examples and membership degrees to (crisp or fuzzy) class labels of these neighbours [60].
- Discernibility Classifier (DC) - uses the discernibility matrix for deriving rules [71].

For the fuzzy rough NN the time taken to build the model is almost 0.01 seconds for each fold which makes it one of the fastest approaches among the algorithms. The average time for building model using Random Forest for each fold is between 0.19 to 0.25 seconds approximately. JRIP is run with minimum total weight of the instances in a rule set to 2, number of optimization runs 2, and no pruning. When we activate pruning the running time for building the model increases drastically; furthermore, the accuracy decrease. Increasing the number of optimization runs also increase the accuracy of the result slightly in some cases. The average run time for JRIP is between 1 to 1.5 second for each fold; therefore, even without pruning this algorithm is slower compared to others. The number of rules generated for each fold varies. Table. 5.6 shows number of the rules generated for each fold in RSES compared to JRIP.

Figure 5.1 gives the prediction of unknown cases with RSES (LEM2 classifier). Here we tested with samples with the class value removed. The coverage parameter for LEM2 was set to 0.99 for training.

Tables 5.7 and 5.8 give the paired t-test results in terms of accuracy for the four classifiers for the two sets of experiments. The following parameters were used: two-tailed, $\alpha = 0.05$ and $n - 1 = 9$ relative to 10 different training-testing runs, where n is the degree of freedom, where α is the significance level and probability distribution (Pr) value = 2.262.

From the results in Table 5.1, the Rough Sets classifier (with LEM2 method) gives the best overall result in terms of overall classification accuracy, accuracy per class,

Table 5.6: Number of rules for JRIP and RSES for both datasets

Fold number	RSES rules 2-status	RSES rules 3-status	JRIP rules 2-status	JRIP rules 3-sta
1	2851	2770	129	226
2	2850	2776	161	186
3	2879	2770	156	187
4	2792	2703	171	204
5	2822	2727	158	198
6	2842	2756	155	222
7	2876	2798	176	227
8	2853	2801	171	211
9	2821	2737	137	209
10	2849	2749	158	212
Mean	2843.5	2758.7	157.2	208.2

```
TABLE testset
ATTRIBUTES 9
 date symbolic, "lab type" symbolic, "lab value" numeric 0,   "lab val cat" symbolic , "sex id"numeric 0,
age numeric 0, diagnosis numeric 0, cbd numeric 0, status symbolic
Objects 15
"2020-08-28 0:00" PSQI 8 2 53 8 3 better
"2020-12-16 0:00" PHQ-9 3 2 71 4 3 better
"2020-08-21 0:00" GAD-7 18 2 63 0 3 better
"2020-11-26 0:00" PHQ-9 11 2 63 2 3 better
"2020-08-25 0:00" PHQ-9 22 2 44 0 1 better
"2020-06-17 0:00" GAD-7 18 2 63 0 3 noChange
"2020-10-07 0:00" GAD-7 18 2 63 3 3 noChange
"2019-04-22 0:00" PHQ-9 2 2 55 0 3 noChange
"2019-09-24 0:00" GAD-7 0 1 68 2 3 noChange
"2019-09-16 0:00" GAD-7 0 2 60 4 3 noChange
"2019-07-07 0:00" PSQI 9 2 60 2 3 noChange
"2020-04-03 0:00" GAD-7 2 1 65 4 3 worse
"2020-10-02 0:00" PSQI 14 1 42 0 3 worse
"2020-12-21 0:00" GAD-7 9 1 65 8 3 worse
"2020-09-10 0:00" PHQ-9 18 2 53 3 3 MISSING
```

Figure 5.1: Prediction results of unknown cases with Rough Sets

Table 5.7: T-test results: Comparison of classification accuracies for Binary Classification

Pairs	Average difference	T-stat
Fuzzy Rough NN Random Forest	0.894	0.002
Fuzzy Rough NN - JRIP	0.048	0.900
Fuzzy Rough NN- Rough Set	2.091	1.717
Random Forest - JRIP	0.942	0.002
Random Forest Rough Set	2.985	6.174
JRIP Rough Set	2.043	5.855

Table 5.8: T-test results: Comparison of classification accuracies for Ternary Classification

Pairs	Average difference	T-stat
Fuzzy Rough NN Random Forest	0.977	0.000
Fuzzy Rough NN - JRIP	0.277	0.381
Fuzzy Rough NN- Rough Set	2.546	5.198
Random Forest - JRIP	0.700	0.086
Random Forest Rough Set	3.523	6.996
JRIP Rough Set	2.823	4.881

sensitivity and specificity values for both the 2-class and the 3-class experiments. The best accuracy (99.34%) was obtained in the ternary classification. It is important to note in both cases (binary and ternary), the classification accuracy is

over 99%. The class distribution in both experiments are highly imbalanced with the Better class having almost 2.5 times more records than the Worse class and 7.9 times the NoChange class. The per class accuracy results are also consistently better in the ternary classification across all three classes. FRNN and JRIP classifiers are second best in terms of overall accuracy. The parameter settings for FRNN (number of K neigbours) and for JRIP were tuned to get the best results. Overall, FRNN gives the next best results.

In terms of sensitivity (or the true positive rate), the best result for the ternary classification is with the better class (Rough Sets classifier- 99.6%). This is not surprising since there are more training examples for this class. In terms of specificity, the best result for the ternary classification is with the NoChange class (or the true negative rate) is 99.9% which is consistent with the accuracy results for this class. Overall, rule-based methods (RSES and JRIP) seem to do better that tree-based ensemble (Random Forest) and Fuzzy Rough nearest neighbour (FRNN) methods for this dataset. However, the number of rules using the LEM2 method is almost 13 times more than the JRIP classifier. This is also reflected in the time it takes for classification. In terms of execution time, FRNN is the best performing classifier. We have presented classification results of other nearest neighbour implementations using different forms of fuzzy and rough sets (Tables 5.5) where FURIA gives overall classification accuracy of over 95% which is again a rule-based classifier. The next best is the discernibility classifier (DC) which is also a rule-based classifier. In terms of Area under the curve (AUC) which is a good indicator of accuracy in datasets where the classes are imbalanced, we can claim the best performance is related to ternary class for NoChange with value of 99.9%. AUC values are calculated by WEKA and the values in parenthesis demonstrate the AUC values calculated manually using the formula:

$$AUC = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{5.1}$$

One of the objectives of this thesis was to predict whether new patients could potentially experience either an increase or decrease in clinical assessment tool scores. We demonstrate this by presenting results for 15 cases (ternary classification) with the LEM2 classifier in Figure 5.1. There are 8 features and one class feature (suppressed). Those cases that have a class label (Better, Worse, NoChange) were correct when compared with the actual values. It also has a label named MISSING.

This is because the LEM2 method has only training coverage of 99% and may not cover all cases. However, of the cases that were classified, all of them were correct. Based on the results of a paired t-test in Tables 5.7 and 5.8, for the binary classification, there is no difference between FRNN and Rough Set classifier (highlighted in blue). There is also no difference between other classifiers. For the ternary classification, there is a difference between Rough Set classifier and the other classifiers (JRIP, RF and FRNN). The reason for the difference is mainly due to need for parameter tuning with other method. However, with the Rough Set classifier one can obtain interpretable rules.

Chapter 6

Conclusion and Future Work

In this thesis, we have worked on the medical data from the Ekosi Health Clinic which their focus is on delivering an innovative health care using cannabinoid therapy especially for chronic pain, mood and sleep challenges. After extensive pre-processing on the dataset, two labelling methods have been proposed to classify the patients based the score of CAT value to binary and ternary classes and a new class feature was indicative of the patient's progress was added to the dataset. Different rough and rough-fuzzy hybrid including rule-based and nearest neighbor methods were applied as well as a tree-based approach. For making the comparison more accurate, we performed the experiments on the same 10 folds for all the algorithms. The result shows rule-based methods performs well especially LEM2, a form of rough set algorithm, has the best performance compared to other methods. Considering the timing of algorithms, FRNN is the fastest algorithm with comparative results. Genetic algorithm has higher accuracy compared to FRNN and more coverage compared to LEM2; however, it takes approximately 4 minutes to generate the rules for each fold which makes it the slowest algorithm. To fulfill an objective of this work, we can predict the progress of new patients having their condition (ex., diagnosis, cannabinoid medication, gender) with promising accuracy. Practically, this study highlights the need for additional research to further identify predictability, patterns and understand the efficacy and real-world evidence regarding cannabinoids for anxiety, depression, sleep disorders, chronic pain, and arthritis. Due to the constraint of precise milligram cannabinoid dosing not being analyzed in this study, future research which incorporates precision cannabinoid-based medicine dosing data is desirable; as the combination of THC and CBD appears to be most beneficial on GAD-7, PHQ-9, and PSQI scores,

further research on the interaction of these two cannabinoids may lead to new and valuable insights, for the benefit of patients and health care practitioners alike. Another future direction can be working on patients with only one visit. More than 160 patients only visited the Ekosi Health center once during the time which data was gathered. This fact highlights the need of further research on these patients as their number is considerable. Since there is no sequence of data for these patients to find a pattern, increase, or decrease in their CAT values and therefore, the trend of progress, we can apply association rule mining approach, Apriori algorithm, to find frequent itemset and general trend in the dataset and as a result generate rules. One may try to extend this work on other diagnosis as in this thesis the focus was on 13 types of diagnosis. Many diagnosis types were removed (278 types) or combined into one category (112 types).

In our case, many patients had only two or three visits in the data set which was not providing enough history (lack of data) to apply time series approaches. However, having access to more patients and monitoring them during a longer period of time, can enable a new path to analyze the data set and apply time series prediction instead of the status computation method. The more the history of patient's feature values, the more possibility of fluctuation. Therefore, in the future with larger data sets, we can consider other methods such as ARIMA algorithm.

Chapter 7

Appendix

Statistic of data sets in each step of preprocessing: Original dataset received from Ekosi: 541 patients, 32514 records

- 3 types of gender: 10881 records for males (188 patients), 21519 records for females (353 patients), and 114 invalid gender records
- 82 types of age: frequent ones are 68, 58, 77, 59, 39 with respectively 1761, 1323, 1287, 1255, 1194 records
- 80 types of cannabinoid formula (order of frequency: CBD oil, CBD oil (20-30mg/ml) & THC: CBD oil (10L15), THC: CBD oil, THC: CBD oils, Yellow oil - CBD 20mg/ml THC $<$ 1 mg/ml)
- 482 types of diagnosis (in order of frequency anxiety, arthritis, insomnia, chronic pain, depressive)

After removing invalid values and separating them into 2 datasets: 375 patients, 29322 records

- 2 types of gender: 10092 records for males, 19230 records for females
- 77 types of age: frequent ones are 68, 58, 77, 59, 39 with respectively 1677, 1281, 1275, 1231, 1110 records
- 75 types of cannabinoid formula
- 390 types of diagnosis

After removing diagnosis and cannabinoid formula not interested: 354 patients, 20026 records

- 2 types of gender: 6854 records for males, 13172 records for female
- 77 types of age
- 73 types of cannabinoid formula
- 112 types of diagnosis

After removing duplicates, the final dataset used for experiments: 354 patients, 8281 records

- 2 types of gender: 2911 records for males (126 male), 5370 records for females (228 female)
- 76 types of age
- 4 types of cannabinoid formula
- 13 types of diagnosis

Table 7.1: Result of Genetic Algorithm for 10-fold cross validation

Fold number	Accuracy (Binary)	Accuracy (Ternary)
1	0.984	0.969
2	0.989	0.982
3	0.996	0.99
4	0.986	0.981
5	0.992	0.983
6	0.994	0.984
7	0.984	0.986
8	0.987	0.978
9	0.99	0.981
10	0.988	0.978
Mean	0.989	0.981

Table 7.1 shows the results of applying Genetic Algorithm option in RSES. Table 7.2 shows the results of FRNN algorithm using option 1 given in Chapter. 4.

Table 7.2: 10-fold cross validation accuracy for FRNN (option 1)

Fold number	Accuracy (Binary)	Accuracy (Ternary)
1	97.22	96.86
2	97.34	96.98
3	96.85	96.37
4	96.85	96.61
5	96.73	97.34
6	97.1	96.37
7	96.61	96.73
8	95.65	96.13
9	97.46	96.73
10	97.1	96.49
Mean	96.89	96.66

Bibliography

[1] S. Ramanna, J. F. Peters, and C. Sengoz, "Application of tolerance rough sets in structured and unstructured text categorization: a survey," in *Thriving Rough Sets*, pp. 119–138, Springer, 2017.

[2] M. E. Lynch and F. Campbell, "Cannabinoids for treatment of chronic non-cancer pain; a systematic review of randomized trials," *British journal of clinical pharmacology*, vol. 72, no. 5, pp. 735–744, 2011.

[3] G. L. Hay, S. J. Baracz, N. A. Everett, P. A. Roberts, J.and Costa, J. C. Arnold, and J. L. Cornish, "Cannabidiol treatment reduces the motivation to self-administer methamphetamine and methamphetamine-primed relapse in rats," *Journal of Psychopharmacology*, vol. 32, no. 12, pp. 415–420, 2018.

[4] P. M., R. Cata, and D. Jutras-Aswad, "Cannabidiol as an intervention for addictive behaviors: a systematic review of the evidence," *SART-S25081*, vol. 9, 2015.

[5] J. L. Lee, L. J. Bertoglio, F. S. Guimarães, and C. W. Stevenson, "Cannabidiol regulation of emotion and emotional memory processing: relevance for treating anxietyrelated and substance abuse disorders," *British journal of pharmacology*, vol. 174, no. 19, pp. 3242–3256, 2017.

[6] P. McGuire, P.and Robson, D. Cubala, W. J.and Vasile, P. D. Morrison, R. Barron, and S. Wright, "Cannabidiol (CBD) as an adjunctive therapy in schizophrenia: a multicenter randomized controlled trial," *The American journal of psychiatry*, vol. 175, no. 3, pp. 225–231, 2018.

[7] C. M. Morin and J. Carrier, "The acute effects of the COVID-19 pandemic on insomnia and psychological symptoms," *Sleep Medicine*, vol. 77, pp. 346–347, 2021.

[8] S. R. Sznitman, S. Vulfsons, D. Meiri, and G. Weinstein, "Medical cannabis and insomnia in older adults with chronic pain: a cross-sectional study," *BMJ supportive & palliative care*, vol. 10, no. 4, pp. 415–420, 2020.

[9] F. G. Ware MA, "Review article: Sleep, pain and cannabis," *Journal of Sleep Disorders & Therapy*, vol. 04, no. 02, 2015.

[10] K. A. Babson, J. Sottile, and D. Morabito, "Cannabis, Cannabinoids, and Sleep: a Review of the Literature," *Current psychiatry reports*, vol. 19, no. 4, 2017. https://doi.org/10.1007/s11920-017-0775-9.

[11] J. Ramesh, N. Keeran, A. Sagahyroon, and F. Aloul, "Towards validating the effectiveness of obstructive sleep apnea classification from electronic health records using machine learning," in *Healthcare*, vol. 9, p. 1450, Multidisciplinary Digital Publishing Institute, 2021.

[12] S. Tariq, N. Akhtar, H. Afzal, S. Khalid, M. R. Mufti, S. Hussain, A. Habib, and G. Ahmad, "A novel co-training-based approach for the classification of mental illnesses using social media posts," *IEEE Access*, vol. 7, pp. 166165–166172, 2019.

[13] J. Kim, D. Lee, and E. Park, "Machine learning for mental health in social media: Bibliometric study," *J Med Internet Res*, vol. 23, p. e24870, Mar 2021.

[14] M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, S. A. Mamun, and M. Mahmud, "Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of alzheimers disease, parkinsons disease and schizophrenia," *Brain informatics*, vol. 7, no. 1, pp. 1–21, 2020.

[15] B. Xiao, C. Huang, Z. E. Imel, D. C. Atkins, P. Georgiou, and S. S. Narayanan, "A technology prototype system for rating therapist empathy from audio recordings in addiction counseling," *PeerJ Computer Science*, vol. 2, p. e59, 2016.

[16] A. B. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine learning in mental health: a scoping review of methods and applications," *Psychological medicine*, vol. 49, no. 9, pp. 1426–1448, 2019.

[17] S. Chancellor, E. P. Baumer, and M. De Choudhury, "Who is the" human" in human-centered machine learning: The case of predicting mental health from social media," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 1–32, 2019.

[18] S. Chancellor, M. L. Birnbaum, E. D. Caine, V. M. Silenzio, and M. De Choudhury, "A taxonomy of ethical tensions in inferring mental health states from social media," in *Proceedings of the conference on fairness, accountability, and transparency*, pp. 79–88, 2019.

[19] O. Edo-Osagie, B. De La Iglesia, I. Lake, and O. Edeghere, "A scoping review of the use of twitter for public health research," *Computers in biology and medicine*, vol. 122, p. 103770, 2020.

[20] A. Sharma and W. J. Verbeke, "Improving diagnosis of depression with xgboost machine learning model and a large biomarkers dutch dataset (n= 11,081)," *Frontiers in big Data*, vol. 3, p. 15, 2020.

[21] M. L. Dobias, M. B. Sugarman, M. C. Mullarkey, and J. L. Schleider, "Predicting mental health treatment access among adolescents with elevated depressive symptoms: Machine learning approaches," *Administration and Policy in Mental Health and Mental Health Services Research*, vol. 49, no. 1, pp. 88–103, 2022.

[22] M. D. Nemesure, M. V. Heinz, R. Huang, and N. C. Jacobson, "Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence," *Scientific reports*, vol. 11, no. 1, pp. 1–9, 2021.

[23] J. B. Edgcomb and B. Zima, "Machine Learning, Natural Language Processing, and the Electronic Health Record: Innovations in Mental Health Services Research," *Psychiatric services*, vol. 70, no. 4, pp. 346–349, 2019.

[24] R. A. Rahman, K. Omar, S. A. Mohd Noah, M. S. N. M. Danuri, and M. A. Al-Garadi, "Application of machine learning methods in mental health detection: A systematic review," *IEEE Access*, vol. 8, pp. 183952–183964, 2020.

[25] C. Su, Z. Xu, J. Pathak, and F. Wang, "Deep learning in mental health outcome research: a scoping review," *Translational Psychiatry*, vol. 10, no. 1, pp. 1–26, 2020.

[26] Z. Pawlak, "Rough sets: Theoretical aspects of reasoning about data. 1991," *Dordrecht & Boston: Kluwer Academic Publishers*, 1991.

[27] Z. Pawlak and A. Skowron, "Rudiments of rough sets," *Information sciences*, vol. 177, no. 1, pp. 3–27, 2007.

[28] N. Shusaku and Y. Kudo, "An application of rough set analysis to a psycho-physiological study - assessing the relation between psychological scale and immunological biomarker," *JACIII*, vol. 13, pp. 352–359, 07 2009.

[29] S. Nomura, H. Santoso, and Y. Kudo, "Data mining the relationship between psychological and physiological stress measurement using rough set analysis," *2010 IEEE International Conference on Systems, Man and Cybernetics*, pp. 3622–3627, 2010.

[30] T. Liu, Y. Chen, Z. Zheng, C. Wang, Z. Hou, C. Chen, and J. Chou, "Using evolutionary rough sets on stress prediction model by biomedical signal," in *International Conference on Machine Learning and Cybernetics, ICMLC 2011, Guilin, China, July 10-13, 2011, Proceedings*, pp. 319–323, IEEE, 2011.

[31] T.-K. Liu, Y.-P. Chen, Z.-Y. Hou, C.-C. Wang, and J.-H. Chou, "Noninvasive evaluation of mental stress using by a refined rough set technique based on biomedical signals," *Artificial intelligence in medicine*, vol. 61, p. 97103, June 2014.

[32] T. Mittal, P. Gupta, and S. Chakraverty, "Application of rough sets in diagnosis of the depressive state of mind," in *2014 Recent Advances in Engineering and Computational Sciences (RAECS)*, pp. 1–6, 2014.

[33] S. Ramanna, N. Ashrafi, E. Loster, K. Debroni, and S. Turner, "Rough-set based learning methods: A case study to assess the relationship between the clinical delivery of cannabinoid medicine for anxiety, depression, sleep, patterns and predictability," *Frontiers in Artifical Intelligence*, 2022. under review.

[34] A. Thieme, D. Belgrave, and G. Doherty, "Machine learning in mental health: A systematic review of the hci literature to support the development of effective and implementable ml systems," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 27, no. 5, pp. 1–53, 2020.

[35] K. Aafjes-van Doorn, C. Kamsteeg, J. Bate, and M. Aafjes, "A scoping review of machine learning in psychotherapy research," *Psychotherapy Research*, vol. 31, no. 1, pp. 92–116, 2021.

[36] D. Galiatsatos, G. Konstantopoulou, G. Anastassopoulos, M. Nerantzaki, K. Assimakopoulos, and D. Lymberopoulos, "Classification of the most significant psy-

chological symptoms in mental patients with depression using bayesian network," in *Proceedings of the 16th International Conference on Engineering Applications of Neural Networks (INNS)*, pp. 1–8, 2015.

[37] A. E. Tate, R. C. McCabe, H. Larsson, S. Lundström, P. Lichtenstein, and R. Kuja-Halkola, "Predicting mental health problems in adolescence using machine learning techniques," *PloS one*, vol. 15, no. 4, p. e0230389, 2020.

[38] M. Srividya, S. Mohanavalli, and N. Bhalaji, "Behavioral modeling for mental health using machine learning algorithms," *Journal of medical systems*, vol. 42, no. 5, pp. 1–12, 2018.

[39] J. Geraci, P. Wilansky, V. de Luca, A. Roy, J. L. Kennedy, and J. Strauss, "Applying deep neural networks to unstructured text notes in electronic medical records for phenotyping youth depression," *Evidence-based mental health*, vol. 20, no. 3, pp. 83–87, 2017.

[40] T. Pham, T. Tran, D. Phung, and S. Venkatesh, "Predicting healthcare trajectories from medical records: A deep learning approach," *Journal of biomedical informatics*, vol. 69, pp. 218–229, 2017.

[41] W. Feng, N. Van Halm-Lutterodt, H. Tang, A. Mecum, M. Mesregah, Y. Ma, H. Li, F. Zhang, Z. Wu, E. Yao, and X. Guo, "Automated mri-based deep learning model for detection of alzheimers disease process," *International Journal of Neural Systems*, vol. 30, p. 2050032, 06 2020.

[42] J. W. Smoller, "The use of electronic health records for psychiatric phenotyping and genomics," *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, vol. 177, no. 7, pp. 601–612, 2018.

[43] N. M. Christ, J. D. Elhai, C. N. Forbes, K. L. Gratz, and M. T. Tull, "A machine learning approach to modeling ptsd and difficulties in emotion regulation," *Psychiatry Research*, vol. 297, p. 113712, 2021.

[44] J. A. Rubel, S. Zilcha-Mano, J. Giesemann, J. Prinz, and W. Lutz, "Predicting personalized process-outcome associations in psychotherapy using machine learning approachesa demonstration," *Psychotherapy Research*, vol. 30, no. 3, pp. 300–309, 2020.

[45] M. Hoogendoorn, T. Berger, A. Schulz, T. Stolz, and P. Szolovits, "Predicting social anxiety treatment outcome based on therapeutic email conversations," *IEEE journal of biomedical and health informatics*, vol. 21, no. 5, pp. 1449–1459, 2016.

[46] T. Althoff, K. Clark, and J. Leskovec, "Large-scale analysis of counseling conversations: An application of natural language processing to mental health," *Transactions of the Association for Computational Linguistics*, vol. 4, pp. 463–476, 2016.

[47] O. Tymofiyeva, J. P. Yuan, C.-Y. Huang, C. G. Connolly, E. H. Blom, D. Xu, and T. T. Yang, "Application of machine learning to structural connectome to predict symptom reduction in depressed adolescents with cognitive behavioral therapy (cbt)," *NeuroImage: Clinical*, vol. 23, p. 101914, 2019.

[48] E. Tolmeijer, V. Kumari, E. Peters, S. C. Williams, and L. Mason, "Using fmri and machine learning to predict symptom improvement following cognitive behavioural therapy for psychosis," *NeuroImage: Clinical*, vol. 20, pp. 1053–1061, 2018.

[49] M. M. Schmitgen, I. Niedtfeld, R. Schmitt, F. Mancke, D. Winter, C. Schmahl, and S. C. Herpertz, "Individualized treatment response prediction of dialectical behavior therapy for borderline personality disorder using multimodal magnetic resonance imaging," *Brain and behavior*, vol. 9, no. 9, p. e01384, 2019.

[50] S. Sivaranjini and C. Sujatha, "Deep learning based diagnosis of parkinsons disease using convolutional neural network," *Multimedia tools and applications*, vol. 79, no. 21, pp. 15467–15479, 2020.

[51] Z. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[52] Z. Pawlak and A. Skowron, "Rough membership functions," in *Advances in the Dempster-Shafer theory of evidence*, pp. 251–271, 1994.

[53] J. Stepaniuk, *Approximation Spaces, Reducts and Representatives*, pp. 109–126. Heidelberg: Physica-Verlag HD, 1998.

[54] L. Zadeh, "Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets Systems*, vol. 177, no. 19, pp. 111–127, 1997.

[55] A. Nakamura, "Fuzzy rough sets," *Note on Multiple-Valued Logic in Japan*, vol. 9, no. 8, pp. 1–8, 1988.

[56] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets*," *International Journal of General System*, vol. 17, no. 2-3, pp. 191–209, 1990.

[57] F. Del Cerro and H. Prade, "Rough sets, twofold fuzzy sets and modal logic fuzziness in indiscernibility and partial information," *The Mathematics of Fuzzy Systems*, pp. 103–120, 1986.

[58] C. Cornelis, M. De Cock, and A. M. Radzikowska, *Fuzzy Rough Sets: From Theory into Practice*, pp. 533–552. John Wiley & Sons, Ltd, 2008.

[59] R. Jensen and C. Cornelis, "Fuzzy-rough nearest neighbour classification," *Transactions on rough sets XIII*, pp. 56–72, 2011.

[60] J. Keller, M. Gray, and J. Givens, "A fuzzy k-nearest neighbor algorithm.," *IEEE Trans. Systems Man Cybernet.*, vol. 15, no. 4, pp. 580–585, 1985.

[61] J. W. Grzymala-Busse, "Rule induction," in *Data mining and knowledge discovery handbook*, pp. 277–294, Springer, 2005.

[62] L. S. Riza, A. Janusz, C. Bergmeir, C. Cornelis, F. Herrera, D. Śle, J. M. Benítez, *et al.*, "Implementing algorithms of rough set theory and fuzzy rough set theory in the r package roughsets," *Information sciences*, vol. 287, pp. 68–89, 2014.

[63] O. Leifler, *Comparison of lem2 and a dynamic reduct classification algorithm*. PhD thesis, Masters thesis, 2002.

[64] S. Sonawani and D. Mukhopadhyay, "A decision tree approach to classify web services using quality parameters," *arXiv preprint arXiv:1311.6240*, 2013.

[65] A. Gupta, A. Mohammad, S. Ali, and M. N. Halgamuge, "A comparative study of classification algorithms using data mining: crime and accidents in denver city the usa," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 7, 2016.

[66] J. G. Bazan and M. Szczuka, "The Rough Set Exploration System," *Springer Transactions on Rough Sets III,*, vol. LNCS 3400, pp. 37–56, 2005.

[67] J. Grzymała-Busse, "Lers-a system for learing from examples based on rough sets," *Intelligent Decision Support*, pp. 3–18, 1992.

[68] R. Jensen and C. Cornelis, "Fuzzy-rough nearest neighbour classification and prediction," *Theoretical Computer Science*, vol. 412, pp. 5871–5884, 2011.

[69] C. Cornelis, M. De Cock, and A. M. Radzikowska, "Vaguely quantified rough sets," in *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing* (A. An, J. Stefanowski, S. Ramanna, C. J. Butz, W. Pedrycz, and G. Wang, eds.), (Berlin, Heidelberg), pp. 87–94, Springer Berlin Heidelberg, 2007.

[70] W. Cohen, "Fast effective rule induction," in *Proceedings of the 12th International Conference on Machine Learning* (A. Prieditis and S. Russell, eds.), (Massachusetts), pp. 115–123, Morgan Kaufmann, 1995.

[71] L. S. Riza, A. Janusz, C. Bergmeir, C. Cornelis, F. Herrera, D. Slezak, and J. M. Benítez, "Implementing algorithms of rough set theory and fuzzy rough set theory in the r package "roughsets"," *Information Sciences*, vol. 287, pp. 68–89, 2014.