Arab American University

Faculty of Graduate Studies

Reformulation of Clinical Depression Symptoms Using Semantically Enhanced Multilingual Lexical Network

By

Fatima Kamel Jamil Khaleel

Supervisor

Dr. Mohammed A. M. Maree

Co-Supervisor

Dr. Mohammad M. Herzallah

This thesis was submitted in partial fulfillment of the requirements for

The Master's degree in

Data Science and Business Analytics

June/2022

Thesis Approval

Reformulation of Clinical Depression Symptoms Using Semantically Enhanced Multilingual Lexical Network

By

Fatima Kamel Jamil Khaleel

This thesis was defended successfully on 25/06/2022 and approved by:

Committee Members	Signature
1. Supervisor Name: Dr. Mohammed A. M. Maree	
2. Co-supervisor Name: Dr. Mohammad M. Herzallah	Digitally signed by Mohammad Herzallah Date: 2022.08.19 09:34:46 -04'00'
3. Internal Examiner Name: Rashid Jayousi	
4. External Examiner Name: Mujahed Eleyat	

Acknowledgments

Without the help of numerous people, this endeavor would not have been feasible. I'd like to convey my heartfelt gratitude to my supervisors, Dr. Mohammed A. M. Maree and Dr. Mohammad M. Herzallah, for their invaluable assistance and input throughout this research. Thanks also to all of my AAUP instructors, friends, and colleagues, without whom I would not have completed my master's degrees.

And my biggest thanks to my family for putting up with the stresses of the past two years, and for their invaluable encouragement and support.

Abstract

Clinical Depression screening is one of the biggest challenges facing psychiatrists, due to the long costly screening process and mental illness-related stigma that creates a serious barrier to mental healthcare. As per World Health Organization (WHO), clinical depression is widely spread and impacts over 264 million individuals around the world and has a large effect on the annual global economy as the lost productivity is estimated at $1 trillion. Therefore, the utilization of approved tested scales in screening, diagnosis, and monitoring response to treatment is important to provide effective needed care. Among the available self-assessment instruments, the Beck depression inventory-II (BDI-II) is one of the most extensively used screening questionnaires for clinical depression. BDI-II, which has 21 categories, shows reliability in discriminating between depressed and non-depressed subjects and identifying the depression severity. The current BDI-II questions' categories were defined in 1996 and have been updated by the DSM-V diagnostic criteria for diagnosing mental health disorders. BDI-II categories which have not been changed since then, are not statistically well divided into independent clusters as they are based on semantically subjective and correlated measures. These non-structured techniques have relatively low accuracy.

To help decrease the treatment gap, it's becoming important to use technology to develop a depression screening application based on previously identified questions by psychiatrists. In this context, natural language processing (NLP) techniques can provide an efficient screening tool to predict clinical depression symptoms and their severity among diagnosed patients. In this research project, we aim to utilize NLP techniques for constructing a lexical-semantic network that enriches the BDI-II questionnaire through the exploitation of extrinsic resources, including WordNet, social media, and unified

medical language system (UMLS). The enrichment is represented by a reformulation of the original questions and their order using semantically derived relationships from the exploited extrinsic resources. Accordingly, we developed a bilingual screening tool using Arabic and English languages to prove more accurate screening and severity diagnosis to help provide the required mental and physical care and lead to efficacious therapy. The tool is designed to be used by patients who can score the questionnaires themselves without seeing a doctor, while the output network offers an immediate utility that can be used by any other researchers who are studying clinical depression symptoms and their impact in a patient-centric manner. In comparison with the current question categories, the developed tool aims to make the screening process faster and more reliable as it keeps only semantic types that are appropriate for clinical depression. It also aims to increase the accuracy of the screening questionnaire by re-grouping the questions that express the same medical semantics based on multiple clustering models.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

API – Application Programming Interface

BDI-II - Beck Depression Inventory-II

BoW - Bag of Words

CIDI-SF – The Composite International Diagnostic Interview

DSM – Diagnostic and Statistical Manual of Mental Disorders

eRisk – Early Risk

FACIT – Functional Assessment of chronic illness therapy

GDS – Geriatric Depression Scale

NLM – National Library of Medicine

NLP – Natural Language Processing

NLTK – Natural Language Toolkit

PHQ – Patient Health Questionnaire

POS – Parts of Speech

PROMIS - Patient Reported Outcomes Measurement Information System

PTSD – Post-traumatic Stress Disorder

QOL – Quality of Life

SIQ – Suicidal Ideation Questionnaire

Tf–Idf - Term frequency– Inverse document frequency

UMLS - Unified Medical Language System

WHO - World Health Organization

Chapter 1 - Introduction

1.1 Background

Clinical depression, which is known to be the "common cold" of the outpatients, is one of the significant public health implications. Depressive disorders are complicated neurobiological conditions that are accompanied by a slew of physiological and cognitive problems. Numerous pathological features that are related to clinical depression were identified in the recent decades (Kaltenboeck & Harmer, 2018). Clinical depression is highly prevalent causing other psychological and physical disorders. However, as patients experience different depressive symptoms ranging from few and mild to many and severe, this makes it hard to screen and treat clinical depression as each type requires different treatment approaches (Kanter et al., 2008; Schaeffer & Jolles, 2019). Therefore, psychiatrists must be careful when screening and diagnosing depression; especially that there are mixed features that can be present in multiple disorders. In addition to that, depression usually causes other mental and organic disorders that must be taken into consideration in treatment.

Unlike the normal emotion of depression, clinical depression that highly impacts the patients' quality of life is a serious mental disorder. Individuals with clinical depression exhibit mixed features that can be present in multiple disorders, some of them can be attributed to treatable organic causes. Symptoms of clinical depression vary from mild to severe (Nawshad Farruque, 2020). Clinical depression shares features with multiple disorders, hence, its definitive diagnosis requires expertise of medical professionals using lengthy and manual clinical interviews (Douglas M. Maurer, 2018; Maurer et al., 2018).

Taking a closer glance at the Palestinian community, it was found that the most frequent psychiatric disorder comorbid with posttraumatic stress disorder (PTSD) is major depression (Madianos et al., 2011), where PTSD develops in some people after experiencing a psychological trauma. [1] Studies show that following the second Intifada of 2000-2007, 68.9% of Palestinian adolescents in Gaza Strip had PTSD and 40% suffered moderate or severe depression (Elbedour et al., 2007). With clinical depression prevalence of more than 24% in the Palestinian community, (Madianos et al., 2012), it is of utmost significance to develop automated tools for clinical depression screening to increase screening rates and identify people with mental health concerns. As these high numbers confirm that early detection and treatment of clinical depression is essential in reducing the emotional burden of the disease, in this project we aim to introduce a network that helps in screening for clinical depression on one hand, and link clinical depression symptoms to other semantically relevant disorders on the other.

In addition to the variety of symptoms caused by clinical depression. Depressive disorders also incur significant financial costs, including time away from work and increased health care costs. The overall cost of depressive disorders that is annually spent in the United States could surpass $40 billion. Suicide is the greatest to pay (Kanter et al., 2008).

Depression screening is the essential for early detection, diagnosis, and treatment. Depression screening for adults is advised, and it should be done using proper methods to guarantee accurate diagnosis, successful treatment, and timely follow-up. All people above the age of 18 should be screened regardless of their risk factors (Douglas M. Maurer, 2018).

[1] https://www.nimh.nih.gov/health/topics/post-traumatic-stress-disorder-ptsd#:~:text=Post%2Dtraumatic%20stress%20disorder%20(PTSD,danger%20or%20to%20avoid%20it.

For depression screening, standardized, self-assessment screening questionnaires are widely used (Sheehan & McGee, 2013). Since 1990, computerized screening for depression has been available, with the potential large-scale population screening over the internet. (Cronly et al., 2018). However, in some countries and languages, one of the difficulties challenging depression screening is the lack of proper tools for objective and systematic evaluation of depression (Kavasis et al., 2005). Many of the regularly used tools for diagnosing clinical depression were developed prior to the fourth edition of the diagnostic and statistical manual of mental illness (DSM-IV), while others were modified to better cover the DSM-IV criteria (Beck et al., 1996). The lack of medical knowledge resources and tools makes it difficult for healthcare professionals to screen for clinical depression. As we are targeting the Palestinian society, we have focused our efforts on clinical depression screening tools in Arabic language as well as English language.

Despite the health burden of depression, it largely goes undiagnosed and undertreated. Previous studies have shown that only 50% of people with diagnostic depression level have received medication, therapy, or both. Also, among those who have received treatment, only 25% are usually treated according to the recommended treatment guidelines. Young adults are more likely to go undiagnosed and undertreated (Arnaez et al., 2020).

Lack of perceived need for treatment, stigma, pessimism about treatment effectiveness, lack of access to treatment owing to financial constraints, as well as other structural barriers such as difficulty or inability to get an appointment are all thought to be impediments to getting the appropriate mental health (Mojtabai et al., 2011). Many people are unable to get the required treatment or seek appropriate healthcare due to these and other issues. However, stigma is usually considered to be the prominent and influential barrier towards getting treatment. When examining stigma, there are two forms to explore: perceived stigma that is represented by the individual's perceptions about the public's attitudes, and the internalized stigma that is represented by how an individual applies these attitudes when having the stigmatizing condition (Dockery et al., 2015; Gurung et al., 2022). Both types are barriers prevent people from obtaining the needed mental health care. However, internalized stigma was found to be a stronger barrier than perceived stigma (Clement et al., 2015; Corrigan et al., 2014).

Stigma affects people in a variety of ways, including an attitude barrier to obtaining mental health treatment, the annoyance of addressing personal issues, embarrassment over the fact of being in treatment, stereotyping, discrimination by local groups, and reduced willingness to engage in mental health care (Vogel et al., 2011). Attitudes about depression and seeking help are directly affected by cultural background (Brown et al., 2010). In some communities, mental illness can be taken as a sign of an inherent flaw in the family, while in other communities, it tends to view mental illness as a personal flaw as it impedes their ability of being independent and achieving success without assistance (Arnaez et al., 2020; Brown et al., 2010).

Screening tools for depression have flaws such as inaccurate results, the need for repetitive examinations and further diagnostic measurements. Studies show that most screening tools had low quality and applicability. Unlike the currently used depression screening measures; self-administered screening tools should be short, easy to administer, and score (Miller et al., 2021). When choosing a depression screening tool, it's crucial to consider performance as well as convenience of use. However, these factors alone do not guarantee the tool's efficacy. The effectiveness of a tool is determined by whether it is acceptable to the general, and whether the benefits of the screening provided by the tool to outweigh the risks (Colligan et al., 2020). As a result, a highly accurate questionnaire with questions that are divided into well-defined independent clusters is requires. A combination of NLP approaches, medical information extraction and retrieval using extrinsic medical ontologies is presented to address these difficulties.

It is known that NLP and the construction of medical knowledge resources have recently become two important fields of medical artificial intelligence. Multiple semantic resources and classification systems have been implemented in the medical information processing domain during the last few years to solve various issue (Reátegui & Ratté, 2018) including: MetaMap[2], a tool for mapping medical text into unified medical language system (UMLS[3]) which is “a large-scale biomedical thesaurus that provides specifications of biomedical knowledge”.

With the availability of medical information resources and modern NLP techniques, we will offer a novel clinical depression screening tool in this project that will aid healthcare professionals in detecting and treating people who have been diagnosed with depression. To achieve this, we will integrate multiple NLP-based pipelines and medical semantic resources to create a lexical semantic network that provides an analysis of the relationships between the questions of Beck depression inventory-II (BDI-II questionnaire). This will facilitate the identification of organic causes of secondary clinical depression and will make the screening process more precise and efficient.

[2] https://metamap.nlm.nih.gov/
[3] https://www.nlm.nih.gov/research/umls/index.html

The following is the structure of the rest of the chapter: the motivation behind this project, a review of the literature and the main elements and characteristics of the proposed system, research methodology, contributions, and thesis structure.

1.2 Motivations

The motivation for implementing this project is the urgent need of a fast and highly accurate clinical depression measurement tool. Such tool helps in providing the necessary treatment and psychological assistance to the patients to reduce depression-related diseases and other side effects that depression may lead to, such as suicide or living an abnormal social and work life. With the proposed enhancement of BDI-II screening tool, we believe that we can contribute to improving the domain of depression screening worldwide and across the Palestinian community by providing a bilingual integrated medical knowledge based self-use screening tool.

1.3 Problem Statement and Research Questions

Depression screening is carried out using manual and computerized measures which suffer from limitations that may lead to incorrect screening. One of the famous and widely used measures is the BDI-II assessment tool. Despite the BDI-validity II's and reliability, the lack of well-established community-based depression screening measures makes early detection problematic (García-Batista et al., 2018). Therefore, in this research project we aim to produce an updated set of reformulated questions in Arabic and English that screen for clinical depression and establish links to other relevant disorders.

We propose the research questions that we will study and address during our research project to reach our goal as follows:

Q1: How are BDI-II questions grouped and what other groupings of questions can be suggested using NLP pipelines?

Q2: What are the relationships between BDI-II questions and existing categories, and what relations can be extracted from extrinsic resources to enhance the questionnaire?

Q3: Will existing medical semantic resources be useful in reformulating the BDI-II questionnaire, and if so, in what contexts (in terms of their conciseness and rigorousness)?

1.4 Research Objectives and Methodology

In this research project, we aim to achieve the following objectives:

1- Designing and developing a bilingual clinical depression screening prototype that exploits lexical semantic information encoded in existing medical knowledge resources.

2- Study the effectiveness of currently used questionnaires for the screening of clinical depression and how they can be improved for more accurate and efficient results.

3- Establish linkages between clinical depression symptoms and other disorders that are not explicitly addressed by the current screening methods/approaches.

To achieve our research objectives; the following steps were carried out during our research work:

1.4.1 NLP Pipeline Construction

In the pre-processing phase, we have applied a multi-level NLP pipeline on BDI-II questions. The pre-processing phase included text cleaning, case-folding, contractions expansion, special characters removal, tokenization, and lemmatization. Based on

WordNet n-grams algorithms, original questions of BDI-II questionnaire didn't consist of any bi or trigrams. Thus, the text was tokenized into single tokens. Then, text was enriched by a semantically related set of terms using UMLS metathesaurus and WordNet to find relations of synonymy, meronymy and hypernymy between the original terms of the BDI-II questionnaire and the expanded set of terms of UMLS. The relationships were then used to build a lexical semantic network (Dancygier, 2017). In addition to the UMLS metathesaurus, e-Risk 2021 dataset[4] was also used to enrich our set of terms by adding the most frequently used terms by people screened with clinical depression on their social media accounts. More details about our NLP pipeline are provided in chapters 3 and 4.

1.4.2 Term Expansion Using Wordnet, UMLS and Other Extrinsic Resources

As our research project aims to reformulate the currently used questions in BDI-II; we have used extrinsic resources for enriching and enhancing the terms used in formulating the questions. For that we have used the e-Risk 2020 dataset which consists of a collection of posts that people diagnosed with clinical depression posted on their social media accounts. In addition to that, we have browsed the UMLS metathesaurus for retrieving semantic relationships between BDI-II questions and other questions in the metathesaurus. This provided us with an expanded set of terms that was then re-expanded using synonymy, meronymy, and hypernymy relationships from WordNet. The final output was then represented using a knowledge graph which was visualized in Python using NetworkX[5] library. Additional details are discussed in Chapters 3 and 4.

[4] https://erisk.irlab.org/2021/index.html

[5] https://networkx.org/

1.4.3 Feature Extraction and Representation

For feature extraction and textual representation for BDI-II text, we used various approaches, such as the Bag of Words, TF-IDF and N-grams. These approaches can handle individual words, but do not reflect the relationship among the words, structure, and semantics.[6] Based on the terms' frequency and the Bag of Words that represents the intersection between BDI-II terms, the UMLS expansion terms and the social media used terms by patients diagnosed by clinical depression. These terms are used to construct a lexical semantic network which represents semantic relations among the terms. The strategies for feature extraction and representation are discussed in depth in Chapters 3 and 4.

1.4.4 Overlapping Questions Detection and Terms' re-weighing

Several overlapping terms developed from the terms' expansion of BDI-II using UMLS and social media. We used a term re-weighting technique to identify questions with the same semantic expression in order to eliminate question redundancy. Weights for these questions are re-assigned based on their frequency and overlap in multiple extrinsic resources. The new weights are taken into consideration for reformulating the existing questions of BDI-II questionnaire, where terms that express the semantic relationship are grouped under the same class. Each class represents a category of questions in the questionnaire. The reformulated questionnaire is then presented to psychiatrists and experts to assess the possibility of using the formulated questions for clinical depression screening. Further details are discussed in Chapters 3 and 4.

[6] https://medium.com/cloudzone/latest-state-of-the-art-models-for-text-representation-feature-extraction-and-text-classification-d821248f0e5c

1.4.5 Topic Modeling and BDI-II Questions Clustering Techniques

Multiple unsupervised machine learning models were used to cluster BDI-II questions into more compact and semantically relevant groups. First, we have developed a lexical semantic network that uses synonymy, hypernymy, meronymy and medical semantic relationships to link n-gram tokens from both the text extracted from questions and the text extracted from medical extrinsic resources. To represent the knowledge graph, NetworkX library in Python was used. Second, we employed multiple clustering techniques including K-Means[7] and TF-IDF[8] using Scikit-learn[9] machine learning library to recategorize similar questions into different groups. The purpose of topic modeling is to show how utilizing of medical knowledge resources can improve the quality of such clustering approaches.

1.4.6 Questions' Reformulation

After identifying relationships between terms in BDI-II questions. An expansion of terms is presented by creating a larger set of terms using extrinsic resources such as WordNet to retrieve terms with synonymy, hypernymy and meronymy relationships, UMLS metathesaurus to retrieve terms with medical semantic relationships and e-Risk 2020 dataset to retrieve terms that are usually used by people that are diagnosed with clinical depression on their social media accounts. The Bag of Words Model is then applied to identify the most frequent terms in each of the datasets as it provides a vectorization of counts of number of occurrences of a word in a text. The most frequent

---

[7] https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a
[8] https://medium.com/@MSalnikov/text-clustering-with-k-means-and-tf-idf-f099bcf95183
[9] https://scikit-learn.org/

terms are then used for questions' reformulation based on these terms' relationships with the original BDI-II questions' terms.

1.4.7 Results Evaluation

Several accuracy metrics are used to evaluate the resulting clusters of the various approaches. We used each of the proposed models to re-categorize items based on the original BDI-II topic and scoring data. We compared the produced models in a comparative analysis. Aside from the comparison analysis, the newly given terms from the extended network are also utilized to improve the terms in the e-Risk 2020 dataset and to investigate the impact of adding these terms on the accuracy of sentiment analysis models used to that dataset. Additional details about the developed models are proposed in Chapters 3 & 4.

1.5 Contributions

We used each of the proposed models to re-categorize items based on the original BDI-II topic and scoring data. We compared the produced models in a comparative analysis. Aside from the comparison analysis, the newly given terms from the extended network are also utilized to improve the terms in the e-Risk 2020 dataset and to investigate the impact of adding these terms on the accuracy of sentiment analysis models used to that dataset.

1.6 Thesis Structure

The following is how the rest of the thesis is organized: We present a background on clinical depression, screening techniques, lexical semantic networks, and medical knowledge sources in Chapter 2. In addition to that, we discussed related studies of existing analysis of the BDI-II questionnaire and other screening tools. In Chapter 3, we provide an overview of our proposed approach as well as an overview of the system architecture, while Chapter 4 provides a full discussion of the approaches used. The experimental work done during this project, as well as an evaluation of the approaches, are presented in Chapter 5. Finally, Chapter 6 discusses the conclusion and future work.

Chapter 2 - Literature Review

2.1 Background

Clinical depression is “the more-severe form of depression, also known as major depression or major depressive disorder”. Clinical depression is a serious mental disorder that has a significant effect on the patients’ quality of life. It affects their work, relationships, and education (Levis et al., 2019). According to data, the number of people suffering from depression has increased by 18 percent in the last decade, with female patients accounting for a far higher percentage than males (Geetanjali Sharma, 2021).

Clinical depression is seen as a neuropsychiatric aliment by the general public and the mainstream media (Ann M. Schaeffer, 2019). Due to multiple reasons, ranging from social stigma to just ignorance; most people who suffer from depression don’t acknowledge it (Nawshad Farruque, 2020). Furthermore, patients might feel depression in a variety of ways. Some people experience a few minor symptoms, while others have a lot of them. Clinical depression is the term used when a patient's symptoms have progressed to the point that they seek professional help.

Each form of clinical depression, and its subypes, require different treatment approaches. Therefore, psychiatrists must be careful when screening and diagnosing depression; taking into consideration that mixed features can be present in multiple diseases. In addition to that, clinical depression usually causes other mental and physical problems that must be taken into consideration in treatment (Savoy & O'Gurek, 2016).

For screening mental disorders, there are professional checklists and screening tools that are used by clinicals and psychiatrists. Some ask direct questions while others are based on a numerical scale to tell whether the person is depressed or not, and the severity of their depression. However, symptoms might be overlooked to be caused by other medical conditions (Brooke Levis, 2019).

There are numerous tests and scales that can be used by clinicians for screening and diagnosing clinical depression, and therefore developing treatment plans (Douglas M. Maurer, 2018). However, the currently used methods are costly, and time consuming. In addition to that, tests' scores might not always be reliable (Guillot-Valdés et al., 2019). Some patients might be embarrassed by their symptoms and might not respond accurately to minimize their score. Other patients might be afraid that the clinician might underestimate their suffering, so they exaggerate with their answers (Eack et al., 2006).

In addition to a variety of screening techniques and approaches, the Beck depression inventory II (BDI-II) is a well-known and commonly used instrument for assessing depression in clinical practice. Recent studies support the validity and reliability of BDI-II (Calvo et al., 2017) (Savoy & O'Gurek, 2016). However, more research is needed to see if the BDI-II can distinguish between depressed and non-depressed people. Furthermore, the BDI-II may be subject to social desirability bias.

2.2 Clinical Depression: Factors and Causes

Clinical depression can strike at any age, but it is more common in adults. It is now recognized that it can affect children and teenagers as well. Other major medical illnesses, such as cancer, diabetes, and heart disease, can coexist with depression. These conditions usually become worse when the patient also suffers from depression. Medications used to treat these aliments can have negative side effects that contribute to depression. Only doctors who have dealt with these types of illnesses before may advise on the best treatment options.[10]

Sometimes there is a trigger for depression. Risk factors include:

- Genetics.
- Brain chemistry.
- Poor nutrition.
- Depression in the medical history of the individual or his/her family.
- Major changes in lifestyle, trauma, or stress.
- Certain medical conditions or drugs.[11]

2.3 Symptoms and Effects of Clinical Depression

Clinical depression symptoms vary depending on the severity of depression. In addition to physical and mental problems, clinical depression may affect the patient's social and work lifestyles. Signs and Symptoms of depression may include:

- Feelings of sadness, tearfulness, emptiness, or hopelessness.
- Anger, frustration, or irritability.
- Loss of interest or pleasure in almost all normal/usual activities, such as sports, hobbies, and sex.
- Disturbances in sleep.
- Tiredness and lack of energy.
- Changes in appetite and losing or gaining weight.
- Anxiety, agitation, or restlessness.
- Feelings of worthlessness or guilt.

[10] https://www.nimh.nih.gov/health/topics/depression
[11] https://www.verywellmind.com/common-causes-of-depression-1066772

- Fixating on past failures or blaming self.
- Persistent thoughts of death, suicidal thoughts, suicidal attempts, or suicide.
- Physical difficulties that are not explained, such as back pain or headaches.[12]

2.4 Screening and Diagnosis of Clinical Depression

All persons over the age of 18 should be screened for clinical depression, regardless of risk factors. Screening for clinical depression can be performed using assessment tools such as commonly used screening questionnaires like patient health questionnaire-2 (PHQ-2), patient health questionnaire-9 (PHQ-9) and Beck depression inventory-II (BDI-II). Measurement tools assess various symptoms of depression. Each question carries a specific numeric value which is then added up to determine if symptoms cross the threshold of clinical significance and its severity. Depression screening tools have been translated into multiple languages and are widely used by psychiatrists and by patients for self-assessment.[13] If the screening is positive, a psychiatrist should confirm the diagnosis using criteria from the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM5). Other mental health diseases should be examined if symptoms do not fulfill the depression diagnosis criteria (Douglas M. Maurer, 2018).

[12] https://www.mayoclinic.org/diseases-conditions/depression/expert-answers/clinical-depression/faq-20057770
[13] https://mshp.mountsinai.org/web/mshp/blog-barriers-to-screening-for-depression

Table 1: Summary of Commonly Used Depression Screening Measures

Measure	Items	Scale	Range	Recall Period	Time to complete
PHQ9[14]	9	4-point	0-27	Last 2 weeks	< 3 minutes
CES-D[15]	20	4-point	0-60	Past week	10 minutes
GDS-15[16]	15	Yes or no	0-15	Past week	2-5 minutes
SCL-90-D[17]	16	5-point	16-80	Past 7 days	< 5 minutes
HADS-D[18]	7	4-point	0-21	Past 7 days	< 3 minutes
BDI-II[19]	21	4-point	0-63	Last 2 weeks	5-10 minutes

2.5 Beck Depression Inventory Screening Tool

The Beck depression inventory (BDI) "is one of the most widely used depression screening tools for measuring the severity of depression in both adults and adolescents over the age of 13"[20]. The inventory is composed of items related to depressive symptoms such as: cognitions, physical symptoms, hopelessness and irritability. The BDI is available in two versions: the original BDI (BDI-IA), which was initially published in 1961 and later revised in 1971, and the BDI-II, which was published in 1996 and was developed to correlate with the DSM[21]-IV criteria for depression.

[14] https://patient.info/doctor/patient-health-questionnaire-phq-9
[15] https://cesd-r.com/
[16] https://patient.info/doctor/geriatric-depression-scale-gds
[17] https://bmcpsychiatry.biomedcentral.com/articles/10.1186/s12888-016-1014-3
[18] https://academic.oup.com/occmed/article/64/5/393/1436876
[19] https://strokengine.ca/en/assessments/beck-depression-inventory-bdi-bdi-ii/
[20] https://www.pearsonclinical.com.au/products/view/39
[21] https://psychiatry.org/psychiatrists/practice/dsm

2.5.1 Features of the Measures for BDI-II

The BDI-II is a 21-item questionnaire that assesses symptoms and attitudes related to depression. The BDI-II replaced four elements from the original BDI (weight loss, body image change, somatic preoccupation, and work difficulty) with four new ones (agitation, worthlessness, concentration difficulty, and loss of energy).

Based on the last two weeks, the respondent must recollect the significance of each statement relative to: sadness, pessimism, past failure, loss of pleasure, guilty feelings, punishment feelings, self-dislike, self-criticalness, suicidal thoughts or wishes, crying, agitation, loss of interest, indecisiveness, worthlessness, loss of energy, changes in sleeping patterns, irritability, changes in appetite, concentration, tiredness or fatigue, and loss of interest in sex[19].

2.5.2 Training, Scoring and Time Required

The BDI was originally intended to be given by a trained interviewer. The BDI, on the other hand, is now widely self-administered due to its short length and ease of use. To understand the questions, patients must be able to grasp spoken or written English and have a reading ability of fifth to sixth grade[22].

Each BDI-II item is scored on a severity scale ranging from 0-3, with a total score ranging from 0-63:

[22] https://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/beck-depression

Table 2: BDI-II Scoring Table

Score	Depression Severity
0-13	None or minimal range depression
14-19	Mild depression
20-28	Moderate depression
29-63	Severe depression

When self-administered, the BDI takes 5-10 minutes to complete, and 15 minutes when administered by an interviewer[23]. The questions are asked over the course of two weeks, including today, rather than the previous week as in the original BDI.

2.5.3 Translations

The BDI measure is available in multiple languages. The questionnaire was translated into the following languages with validation: Arabic, Chinese, Dutch, Finnish, French, German, Japanese, Persian, Polish, Portuguese, Spanish, Serbo – Croatian, Swedish, and Turkish. While it was translated into the following languages without validation: Cambodian, Italian, Korean, and Xhosa.

2.5.4 Pros and Cons of Using BDI-II[24]

Pros:

1. The BDI-II is commonly used and approved symptomatology as a tool for depression.
2. The BDI-II can be given orally by an examiner to people who have difficulty with reading or concentrating.

---

[23] https://onlinelibrary.wiley.com/doi/epdf/10.1002/art.11410
[24] https://www.nctsn.org/measures/beck-depression-inventory-second-edition

3. The BDI-II is user-friendly; it is easy to administer and score.
4. It has been translated into a variety of languages other than English, and its psychometric properties have been demonstrated in multiple cultural groups including the deaf community.
5. The BDI-II is state-related depression assessment tool that can be used as a short weekly screening tool before therapy sessions.
6. The tool has been demonstrated to be useful in detecting changes in treatment outcome studies.

Cons:

1. Underreporting and overreporting are conceivable due to BDI-II's validity.
2. People with a poor level of education, as well as some Spanish speakers, may struggle with the response structure.
3. The method of determining the cut scores may lead to more false positives or overdiagnoses of depression among people.
4. Some questions ask the respondent to compare their current situation to a previous one (e.g., than usual, as ever). Individuals who have experienced chronic trauma since childhood may circle a zero because they do not feel any worse than "usual".
5. The normative sample is mostly made up of white people (91%).
6. The norms were acquired with adults, even though the measure can be used for teenagers.
7. In the United States, the bulk of psychometric research with teenagers have used largely Caucasian samples and have excluded huge numbers of people from

lower socio-economic backgrounds. More research on the BDI-II with various groups of teenagers is needed.

2.5.5 Arabic Version of BDI-II

The BDI-II Arabic version is a 21-item questionnaire that assesses the intensity of self-reported depression during the previous two weeks. It relates to the DSM-IV criteria for the diagnosis of depressive disorders. The items are scored on a 4-point scale ranging from 0 to 3, similar to the original English edition of the BDI-II. Dr. Ghareeb (Ghareeb, 2000) produced the Arabic version, and psychometric properties were tested in 17 Arabic nations. In Arabic countries, they found the BDI-II to have acceptable validity and reliability. In these countries, Alpha Cronbach's coefficients ranged from 0.82 to 0.93 (Alansari, 2005) (Rahat et al., 2012).

2.6 Barriers and Challenges to Screening for Depression

Clinical depression, as well as other mental and psychological diseases suffer from multiple barriers and interventions. These barriers include stigmatization, acceptance, financial burden, lack of awareness, lack of geographical access, religious and sociocultural impacts. These and other barriers prevent the patients from accessing mental health services. In addition to that, gossip and social visibility within the community have been shown to increase stigma and social exclusion of psychiatric patients (Ko et al., 2013).

2.7 Semantic Relations in Natural Language Processing

Natural language processing (NLP) and neural network-based representations have advanced quickly in recent years, particularly in the field of word representations.

Semantic relations represent the meaningful associations between words in different domains. In the medical domain, a semantic relation represents the prevalent relations between medical terms. With the advancement of technology in the medical field, many medical concepts, and the relationships between them have been encoded in medical ontologies that represent controlled vocabularies (Chen et al., 2018).

The investigation of relations between symptoms in the medical field provides a clear understanding of the meaning of these symptoms. Therefore, this facilitates clinical decision making based on the ontological representation. Table 3 below provides a summary of previously discussed approaches for semantically enhanced clustering techniques. In the table, we highlighted the main features of each approach.

Table 3: Semantically Enhanced Clustering Approaches

Approach	Category	Used Techniques
Clustering of semantically enriched short texts (Kozlowski, M. & H. Rybinski, 2019)	1. Neural-based distributional model 2. External knowledge resources	• Bisecting K-means clustering • Suffix Tree Clustering (STC) • Lingo • SenseSearcher algorithm (SnS)
Text Mining in Biomedical Domain with Emphasis on Document Clustering (Renganathan,V. , 2017)	Concept-based clustering	• Knowledge extraction from biomedical literature. • K-Means clustering
Concept embedding based weighting scheme for biomedical text clustering and visualization (Luo, X. & S.	Concept-based clustering	• Concept Extraction • Concept Embedding Generation • Document Representation

Shah,2018)		• K-Means clustering
Clustering of biomedical documents using ontology-based TF-IGM enriched semantic smoothing model for telemedicine applications (Sandhiya, R. & M. Sundarambal 2019)	Ontology-based clustering	• Token and phrase identification using n-gram • Enriched semantic smoothing model • Mesh ontology • K-Means clustering

2.8 The Unified Medical Language System (UMLS)

The United States National Library of Medicine (NLM) has developed the unified medical language system (UMLS) to unify various authoritative biomedical source terminologies into a unified knowledge representation.[25]

The Metathesaurus is a “multi-purpose, multi-lingual vocabulary database that contains information on biomedical and health related concepts, their names, and the relationships among them”. The Metathesaurus is a collection of electronic versions of thesauri, classifications, code sets, and lists of controlled terms used in patient care, billing, public health, statistics, indexing and cataloging biomedical literature, and basic, clinical, and health services research.[2]

The UMLS metathesaurus allows us to contrast knowledge graphs to represent the medical terms in depression screening questionnaires which helps in extending the original terms with additional lexical semantically related terms that are acquired from the used knowledge graph.

[25] http://www.nlm.nih.gov/research/umls/about_umls.html

2.9 Summary

We've covered the basics of clinical depression, its symptoms, factors and causes, screening techniques, and hurdles to depression screening and treatment in this chapter. In addition, we provided an overview of the usage of Natural Language Processing in the medical domain and the representation of semantic relations, as well as the Unified Medical Language System, which would be utilized as the metathesaurus in our proposed strategy to enrich the BDI-II questionnaire. There is a substantial amount of research on clinical depression screening. However, there is a lack of work in the field of textual enrichment utilizing NLP and lexical semantic knowledge graphs for depression screening, and this needs to be addressed to close the gap.

Chapter 3 - System Overview

As depicted by Figure 1, we introduce a high-level overview of the steps of our proposed method. The first input is the text of Beck Depression Inventory Questionnaire (BDI-II). The questionnaire consists of 21 questions aiming to measure the severity of depressive symptoms, assessing the intensity of depression in psychiatric and normal populations, ages 13 and above (Gebrie, 2018). Using Python, text pre-processing and analysis was performed to have a better understanding of the BDI-II questionnaire by analyzing unique terms used in this questionnaire, the labels given to the questions and the weight of the words used in formulating these questions. After having an overview of the input questionnaire; extrinsic databases like WordNet and UMLS Metathesaurus are used to create a lexical semantic network by providing the related semantic relationship between these questions and recategorizing them into new clusters using multiple clustering techniques. The final output is a reformulated depression screening questionnaire that depends on the output lexical semantic network obtained from the relationship network created from UMLS Metathesaurus and BDI-II Questionnaire.

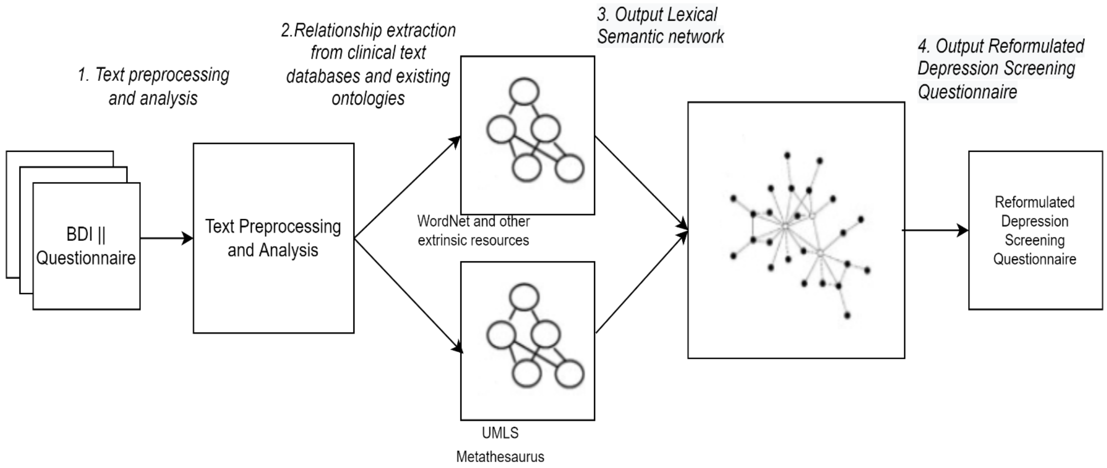

Figure 1: Steps for Reformulating the BDI-II Screening Questionnaire

3.1 System Architecture

In this project, we have constructed a lexical semantic network to represent the relationships linking between BDI-II original terms to their semantically related medical terms that have been deducted from UMLS metathesaurus. To accomplish this task, we have created an expanded Bag-of-Words based on the new terms obtained from the expanded network using UMLS, along with the use of e-Risk dataset to expand our network by terms that are commonly used by people screened with clinical depression on their social media. The new expansion of terms was then used to construct the lexical semantic network using WordNet to identify the relationships between the terms such as synonymy, meronymy and hypernymy. In addition to that, we have utilized multiple pre-trained NLP language models to re-categorize the questions by clustering them into new clusters using multiple clustering techniques such as TF-IDF and K-Means. After identifying the new clusters, we have reformulated the questions by enhancing them using the output BoW that we got after the terms' expansion and questions were re-ordered based on the weight of the words as per their usage in the extrinsic resources used in this project. Figure 2 describes the overall architecture of the proposed techniques. As it shows, the input questionnaire is processed using NLP pipeline that includes text normalization, punctuations removal, tokenization, and lemmatization. The output of the pipeline is then used to enrich the text by expanding these terms and extracting the relationships among them from extrinsic medical and lexical resources including WordNet, e-Risk, and UMLS. The extracted relationships are used to create a lexical semantic network that enriches the terminology of the original questionnaire. This network is used to reformulate the questions based on the weight given to each of the terms in the network depending on their occurrences in the original versus the expanded terms. The questions of the original questionnaire are also re-categorized into new set of categories based on the results of K-Means and TF-IDF clustering models. The final output of the proposed system is a new categorization of the original BDI-II questions after being reformulated and enriched with new terminologies.

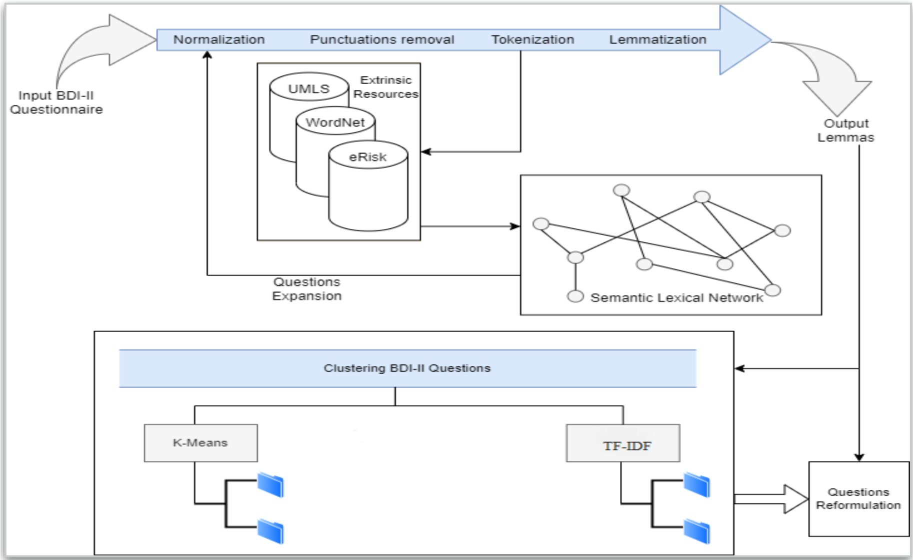

Figure 2: System Architecture

3.2 NLP Pipeline: Step-by-step

The general NLP pipeline normally comprises three main stages: text processing, feature extraction, and modeling. As unstructured data, the input questions of BDI-II questionnaire, was required to go through a multi-level NLP pipeline. During data pre-processing, and as any NLP solution, our solution consists of several components that are chronologically ordered, including handling acronyms and abbreviations, tokenizers, lemmatizers, and others as clarified below:

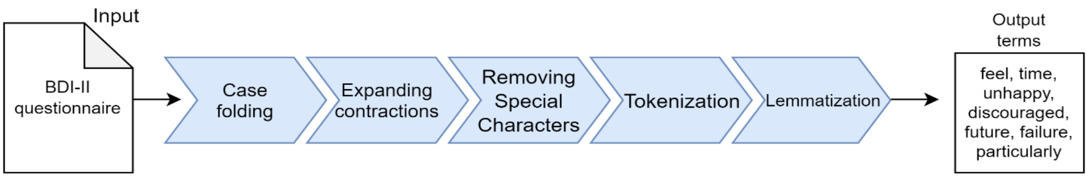

Figure 3: NLP pipeline step-by-step

The following sections provide more details about each of these steps and how they are applied in our project.

3.2.1 Case-folding

Case-folding is a common strategy in an NLP pipeline by which all letters are reduced to lowercase. Case-folding aims to simplify the text and maintain its consistency and reduce the impact of the differences among tokens during the NLP tasks. The simplest heuristic is to convert all words at the beginning of the sentences to lowercase words. However, it is also important to take into consideration that some words must remain capitalized – usually mid-sentence capitalized words –. [26] Case-folding was used as the first step of our NLP pipeline to pre-process BDI-II questions as part of text normalization, by lower casing all of the questionnaire's tokens, regardless of their category, so that superficial differences between tokens have no impact on the quality of clustering techniques.

Before Lowercasing: 0 I do not feel sad. 1 I feel sad much of the time. 2 I am sad all the time. 3 I am so sad or unhappy that I can’t stand it.
After Lowercasing: 0 i do not feel sad. 1 i feel sad much of the time. 2 i am sad all the time. 3 i am so sad or unhappy that i can’t stand it.

[26] https://nlp.stanford.edu/IR-book/html/htmledition/capitalizationcase-folding-1.html

3.2.2 Contraction Expansion

Contraction words or combinations of words that are shortened by deleting letters or substituting them with an apostrophe are called. English Contractions are often created by removing one of the vowels from the word.[27] As part of data pre-processing of BDI-II questions, contractions' expansion is applied to reduce the number of meaningless punctuation and allow better identification of stop words to evaluate their weights among other tokens in BDI-II.

Before Expanding Contractions: 3 i am so sad or unhappy that i **can't** stand it. 0 i **don't** feel i am being punished.
After Expanding Contractions: 3 i am so sad or unhappy that i **cannot** stand it. 0 i **do not** feel i am being punished.

3.2.3 Special Characters Removal

As part of cleaning the textual data of BDI-II, regular expressions (regexes) are used to remove all numeric and non-alphanumeric characters (punctuations and special characters) which add extra noise in the unstructured data. In addition to special characters, white spaces at the beginning and at the end of the string are also removed.[28]

Before Removing Special Characters and white spaces: 1 i have thoughts of killing myself, but i would not carry them out. 2 i would like to kill myself.
After Removing Special Characters and white spaces: i have thoughts of killing myself but i would not carry them out i would like to kill myself

[27] https://www.geeksforgeeks.org/nlp-expand-contractions-in-text-processing/
[28] https://towardsdatascience.com/nlp-building-text-cleanup-and-preprocessing-pipeline-eba4095245a0

3.2.4 Tokenization

Tokenization "is a common task in NLP in which a text is separated into smaller units called tokens. Tokens can be either words, characters or subwords".[29] However, we may utilize the tokenization methods to tokenize into n-grams (word sequences), where n is the number of words we want to capture in each n-gram.[30]

After cleaning BDI-II text of all the characters that may cause noise in later stages, text is then tokenized to allow further analysis on the weight of the tokens and their usage in each of the categories. These tokens are later used to enhance the questionnaire by using them for terms' expansion from multiple extrinsic resources.

Before Tokenization: i am so restless or agitated that i have to keep moving or doing something i have not lost interest in other people or activities i am less interested in other people or things than before
After Tokenization: i, am, so, restless, or, agitated, that, i, have, to, keep, moving, or, doing, something i, have, not, lost, interest, in, other, people, or, activities i, am, less, interested, in, other, people, or, things, than, before

3.2.5 Stop Word Removal

Stop words aren't really crucial in a text. Each programming language has its own set of stop words that must be used. Because stop words are the most prevalent words in any language and don't offer much information to the text, they are usually filtered out. We remove the low-level information from our text by removing stop words, allowing us to focus more on the most crucial information.[7]

[29] analyticsvidhya.com/blog/2020/05/what-is-tokenization-nlp/
[30] https://bookdown.org/Maxine/tidy-text-mining/tokenizing-by-n-gram.html

Stop word removal reduces the size of a dataset. As a result of the reduced amount of tokens engaged in the training, the training time is reduced. However, in some cases stop words are not removed from the text. This usually depends on the type of task we are performing on the data and the goal that we want to achieve.[31]

After identifying the set of tokens that we have in BDI-II questionnaire, stop words removal is then applied on the questionnaire's tokens to emphasize the weight of the rest of the tokens and use them for terms expansion.

Before Stop Words Removal: i, am, so, restless, or, agitated, that, i, have, to, keep, moving, or, doing, something i, have, not, lost, interest, in, other, people, or, activities i, am, less, interested, in, other, people, or, things, than, before
After Stop Words Removal: restless, agitated, keep, moving, something lost, interest, people, activities less, interested, things

3.2.6 Stemming

Stemming is "a natural language processing technique that reduces inflection in words to their root forms such as mapping a group of words to the same stem even if the stem itself is not a valid word in language".[32] As part of text pre-processing, multiple stemming techniques are applied to BDI-II tokens to get the stem of each of the tokens.

[31] https://towardsdatascience.com/text-pre-processing-stop-words-removal-using-different-libraries-f20bac19929a
[32] https://www.datacamp.com/community/tutorials/stemming-lemmatization-python

Porter Stemmer

Porter Stemmer which was invented by Martin Porter in 1980 has five steps of word reduction. The original stemmer, the Porter Stemmer, is noted for its ease of use and speed. Porter Stemmer's output is typically a shorter word with the same basic meaning.[33]

Lancaster Stemmer

Lancaster Stemmer is a simple tool; however it frequently yields excessive stemming results. Stems that have been over-stemming become non-linguistic or nonsensical.[8]

Word	Porter Stemmer	Lancaster Stemmer
feel	feel	feel
time	time	tim
unhappy	unhappi	unhappy
discouraged	discourag	disco
future	futur	fut
failure	failur	fail
particularly	particularli	particul
irritable	irrit	irrit
concentrate	concentr	cont
completely	complet	complet

3.2.7 Lemmatization

Lemmatization, unlike stemming, appropriately lowers inflected words, ensuring that the base word belongs to the language. The root word is called “lemma”.[7] As part of text

[33] https://www.analyticsvidhya.com/blog/2021/11/an-introduction-to-stemming-in-natural-language-processing/

pre-processing, lemmatizing techniques are applied to BDI-II tokens to get the lemma of each of the tokens.

Word	Wordnet Lemmatizer
feel	feel
time	time
unhappy	unhappy
discouraged	discouraged
future	future
failure	failure
particularly	particularly
irritable	irritable
concentrate	concentrate
completely	completely

3.3 Sentiment Analysis

Sentiment Analysis is a method that provides automated extraction of opinions and sentiments from text. Machine learning provides multiple sentiment classifications that are based on the context of documents. However, with machine learning approaches, it is difficult to determine the sentiment of text if the text contains sentiment lexicons with contrast sentiment. (Chong et al., 2014)

Sentiment Analysis, which mainly relies on Natural Language Processing techniques, allows researchers and scientists to examine text from a variety of sources. The most basic use of sentiment analysis is to collect people's thoughts on a certain topic. Sentiment analysis, like stop words, is based on a collection of predetermined terms that describe the writer's impression. People's opinions are divided into three categories: negative, neutral, and favorable (Rajput, 2020).

In BDI-II questionnaire, sentiment analysis helps in identifying the sentiment classes to which each of the tokens belong. This allows better understanding of the nature of the tokens in the questionnaire and how to apply terms' expansion in each of the sentiment classes, especially when comparing to the list of tokens used in eRisk dataset as the terms from that dataset are retrieved after applying several sentiment analysis techniques on the data.

Questions	Sentiment
I do not feel sad.	Neutral
I am sad all the time.	Negative
I get as much pleasure as I ever did from the things I enjoy.	Neutral
I can't get any pleasure from the things I used to enjoy.	Negative

3.4 Co-occurrence Network

"Co-occurrence analysis is simply counting paired data within a collection". Co-occurrence of items in the collection presents an association between them. If pairing happens only once, this means that the association is weak. If the pairing happens many times, this means that the association is strong. (Buzydlowski, 2015). In BDI-II we provide an analysis of co-occurring tokens by analyzing non-stop words occurring in the same questions. This provides us a better understanding of the BDI-II context and the terminologies used to formulate its questions and the associations between them.

i get very little pleasure from the things i used to enjoy.
i feel guilty over many things i have done or should have done.

After stop words removal the co-occurrence network of the tokens in this example is presented as shown in Figure 4. The nodes of the network are represented by the words of the questions while the edges are represented by the association of words occurring in the same question.

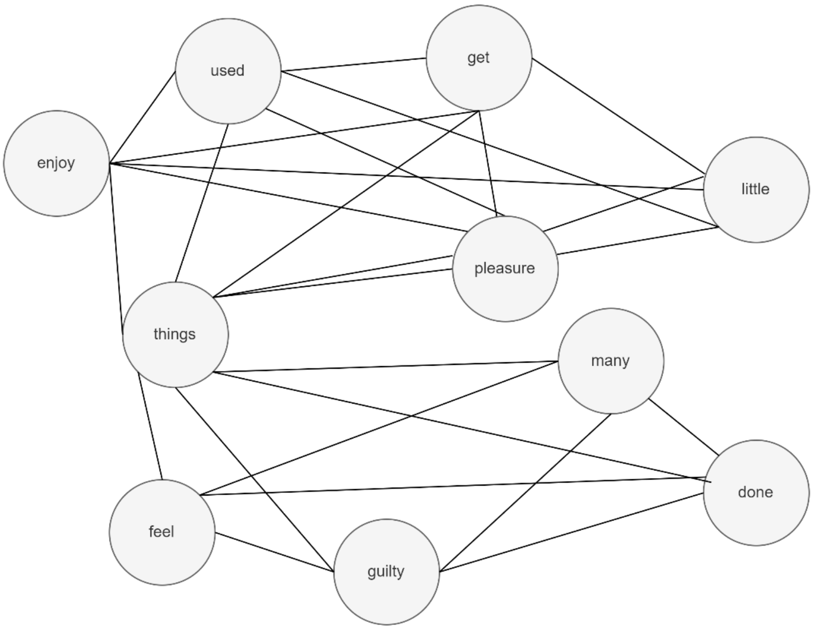

Figure 4: Co-occurrence Network Example

3.5 Clustering Techniques for BDI-II Questions

Clustering is an unsupervised model used to separate similar objects into the same group. Thus, the first step is to identify the number of groups without actually knowing which cluster data belongs to.[34] The original BDI-II questions are clustered into 21 different clusters ordered as follows: Sadness, Pessimism, Past Failure, Loss of Pleasure, Guilty Feelings, Punishment Feelings, Self-Dislike, Self-Criticalness, Suicidal Thoughts or Wishes, Crying, Agitation, Loss of Interest, Indecisiveness, Worthlessness, Loss of Energy, Changes in Sleeping Pattern, Irritability, Changes in Appetite,

[34] https://towardsdatascience.com/making-sense-of-text-clustering-ca649c190b20

Concentration, Tiredness or Fatigue, Loss of Interest in Sex. Multiple clustering approaches were used in our study to determine the number and order of groups in the result questionnaire. For that, we have weighted the terms used in the BDI-II questionnaire, in addition to the list of expansion terms to weigh the terms based on their significance in the context. The following Clustering models were utilized: K-Means clustering, and TF-IDF clustering.

Figure 5 represents the flow in which the questionnaire was processed to identify the new categorization for the questions in BDI-II questionnaire. The text of the questions of BDI-II was processed, and the word embeddings of the text was extracted from multiple resources that are used to enrich the original text. Terms' weighting was then applied to identify the most common terms that were used in certain contexts that were later used for text reformulation. After reformulating the questions, clustering techniques were applied to identify the new categories of BDI-II output questionnaire.

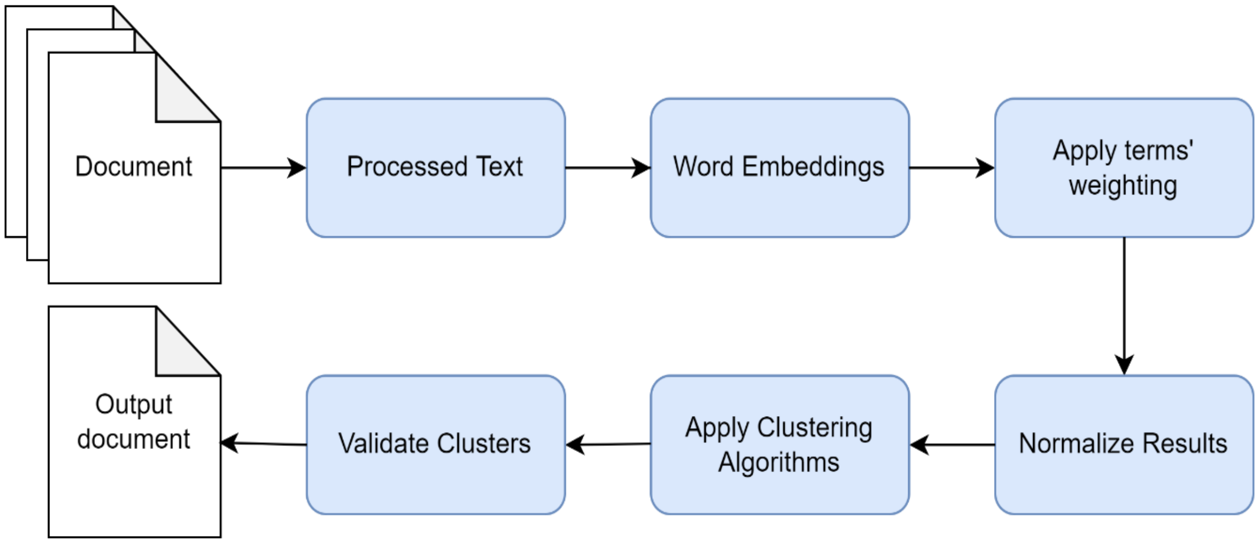

Figure 5: Proposed Re-Categorization Models for BDI-II Questionnaire

3.5.1 Identifying Number of Clusters

In any unsupervised model, there is no conclusive answer to the ideal number of clusters into which the data should be grouped. Some direct methods are used to identify the optimal number of clusters – known as *k* –.[35] In our project we have utilized two of the most popular methods that are: *elbow* and *silhouette*.

Elbow Method

In Elbow, we adjust the number of clusters (k) from 1 to 10. We calculate the Within-Cluster Sum of Square (WCSS) for each k, which is "the sum of squared distances between each point and the cluster's centroid". The WCSS plot looks like an elbow, and we can tell that there is a point that causes the elbow shape by analyzing the graph. The optimal number of clusters is represented by the k value on the x-axis corresponding to this location.[36]

Silhouette Method

The Silhouette approach is used to figure out how far apart the clusters are. The Silhouette plot displays how close each point in one cluster is to points in neighboring clusters, and so aids in determining the appropriate cluster size.[37]

Steps to compute Silhouette Coefficient of a point (i):

1- Compute (i): "The average distance of the point with all other points in the same cluster".

---

[35] https://www.datanovia.com/en/lessons/determining-the-optimal-number-of-clusters-3-must-know-methods/
[36] https://www.analyticsvidhya.com/blog/2021/01/in-depth-intuition-of-k-means-clustering-algorithm-in-machine-learning/
[37] https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html

2- Compute b(i): “The average distance of the point with all points in the closest cluster to its cluster”.

3- Compute s(i): “The silhouette coefficient of the point (i) using the following formula”:

$$s(i) = \frac{b(i) - a(i)}{max(b(i), a(i))}$$

Silhouette coefficient value can be in the range [-1,1] where 1 indicates that the data is close within the same cluster but far from other clusters, and 0 indicates an overlap in the clusters, and -1 indicates that the point is assigned to a wrong cluster.

3.5.2 K-Means Clustering

K-Means is “a famous vector-based clustering algorithm that returns a cluster assignment to one of the k possible clusters for each object”. Defining the distance metric between two data points and defining the number of clusters are crucial in K-Means (Rangrej et al., 2011).

The K-Means algorithm divides the dataset into k non-overlapping subgroups. This means that each point (subject) is clustered into only one group, with intra-cluster points being as similar as feasible to points in other groups and as distinct as possible from points in other groups.[38]

The K-Means algorithm works as follows:

1- Define the number of clusters, known as *k*.

[38] https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a

2- Initialize cluster centroids by shuffling and randomly selecting k data points for each centroid without replacement.

3- Iterating until there are no changes to the centroids to which the data points are assigned.

The K-Means equation is given as follows:

$$J = \sum_{i=1}^{m} \sum_{k=1}^{K} w_{ik} \|x^i - \mu_k\|^2$$

Where $w_{ik} = 1$ for data point xi if it belongs to cluster k; and $w_{ik} = 0$ otherwise.

3.5.3 TF-IDF Clustering

TF-IDF is “the word representation method that can give constant weight to each word”. It denotes the degree to which a word is relevant to a specific document. TF-IDF takes into consideration the frequency of words within a document and the inverse of the frequency of words in the document (Subakti et al., 2022).

The following equation can be used to derive the numerical representation of a word t in a document d by TF-IDF:

$$tf(t,d) = \frac{f_d(t)}{w \in d \, max f_d(w)}$$

3.6 Creating a Lexical Semantic Network

A lexical semantic network is “a type of semantic network which represents the relations between words, sub-words or some-other linguistic related terms”. [39] For creating a lexical semantic network for BDI-II terms, we will use UMLS, eRisk social

[39] https://ai.stackexchange.com/questions/2634/what-exactly-are-the-differences-between-semantic-and-lexical-semantic-networks

media posts and WordNet to create an expanded set of terms and retrieve the relationships among them. A knowledge graph is then created to represent these relationships. The terms are re-weighed based on their occurrences in all 3 resources and a BoW is created for the most common terms in all resources to be used in the questions' reformulation. This BoW is also compared with the google English translation of the unique Arabic tokens that we get from the Arabic Version of BDI questionnaire. The Arabic version is enriched using the BoW we get from the English version of BDI-II and questions are then reformulated accordingly.

3.7 Summary

To summarize, the goal of this chapter is to provide an overview of our proposed system architecture from the NLP pipeline used to pre-process the data, questions' clustering, and lexical semantic network creation. The pipeline consists of multiple phases including case-folding, contractions' expansion, special characters removal, tokenization, stop words removal, stemming, and lemmatization. The pipeline output is then used to enrich the questions by expanding the tokens set using UMLS, eRisk and WordNet to extract relations on both questions and tokens levels. The resulting network is used to create an expansion of tokens that are then used to reformulate the questions of BDI-II. Further details are discussed in Chapter 4.

Chapter 4 - System Implementation Details

This chapter presents the implementation details of the proposed system presented in Chapter 3. This includes the NLP pipeline and data pre-processing, terms' expansion using extrinsic resources, creating the lexical semantic network, terms' re-weighting, and clustering techniques, and finally the reformulation of questions based on the relationships extracted from multiple medical and semantic networks.

To perform this part, we used Python programming language in Google's colab notebooks for implementing the proposed NLP pipeline and utilized it to enrich the BDI-II questions and create a reformulation of the questions (Code can be found at the link in the footnote[40]). The system used for the code implementation is a Lenovo ThinkPad T14 Laptop with an intel® Core i7 8-logical processor and 16 GB RAM.

4.1 NLP Pipeline Development

Data pre-processing includes case-folding, contractions' extraction, text cleaning, tokenization and lemmatization as follows:

1. BDI-II questionnaire was inputted into Python notebook in the form of a csv file that was then imported into a DataFrame using pandas[41] Python library.
2. Case-folding was applied to all questions converting all words into lowercase to reduce significance of uppercase words.
3. Contractions were expanded to identify the original terms and reduce noise of apostrophes as a punctuation.

[40] https://colab.research.google.com/drive/1SQC0wWLcWgyTaB6-6xsa2YkliJtoKClC#scrollTo=eIVPbrqFuK-t
[41] https://pandas.pydata.org/

4. Special characters, numbers and white spaces were removed.
5. Clean data was then tokenized using nltk[42] word tokenizer. Potential n-gram tokens were analyzed using WordNet.
6. Stemming and Lemmatization were performed using Porter and Lancaster Stemmers and WordNet Lemmatizer.

During the pre-processing stage, stop words removal was performed as part of the NLP pipeline. However, the analysis showed that about 62% of the questionnaire tokens are stop words. Therefore, stop words removal isn't relevant in our case.

In addition to that, as a final output for the NLP pipeline, lemmas were considered rather than stems since lemmas return a root that is an actual English word while stems do not always have a meaning.

4.2 BDI-II Questionnaire Exploration and Sentiment Analysis

The original BDI-II questionnaire consists of 21 questions' categories, an overall of 90 questions with numeric values from 0 to 3 used to calculate the assessment score by adding up the score of all the questions to determine whether the subject is depressed or not, and the level of severity.

4.2.1 Data Statistics

In the table below, some statistics about the nature of data in the BDI-II questionnaire is presented as resulted from the data analysis in Python using NLTK tokenizer.

[42] https://www.nltk.org/

Table 4: Data Statistics for BDI-II Questionnaire

Number of categories	21
Number of questions	90
Number of tokens	589
Number of tokens after contractions' expansion	787
Number of unique tokens	166
Number of stop words	482
Average number of characters per token	3.75

The tokens in the questionnaire vary in their length. The average number of characters per token is 3.75 as seen in Figure 6.

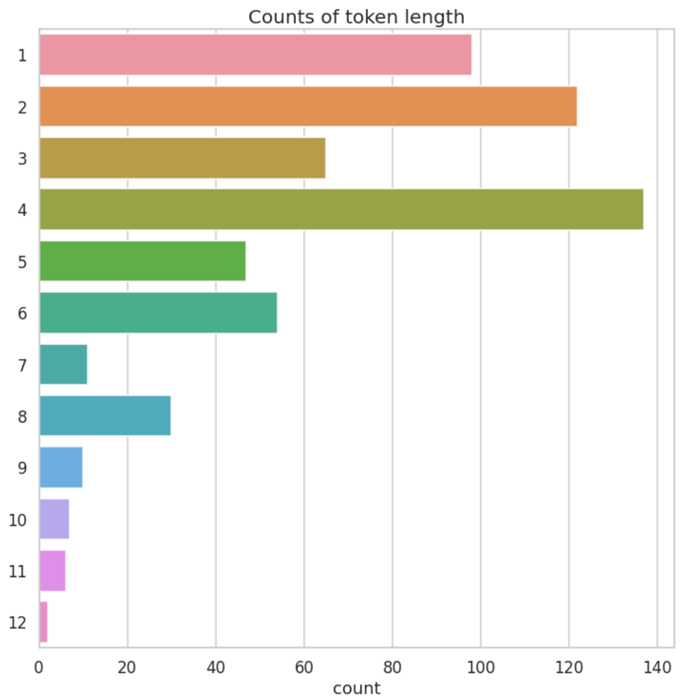

Figure 6: Number of Tokens Based on Tokens' Lengths in BDI-II Questionnaire

After data pre-processing was performed, tokens were analyzed in terms of the most common unigrams and bigrams occurring in the dataset as illustrated by the results depicted in Figures 7

and 8.

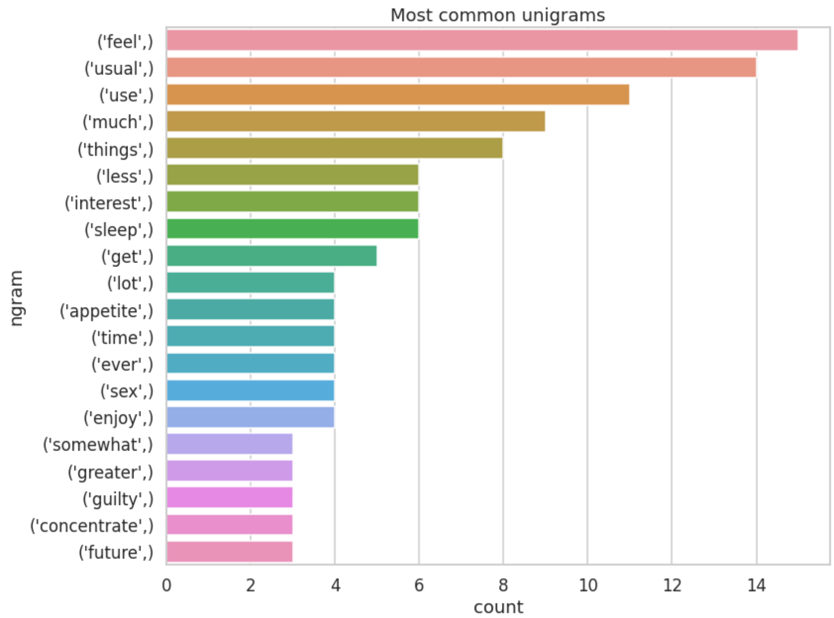

Figure 7: Most Common Unigram Tokens in BDI-II Questionnaire

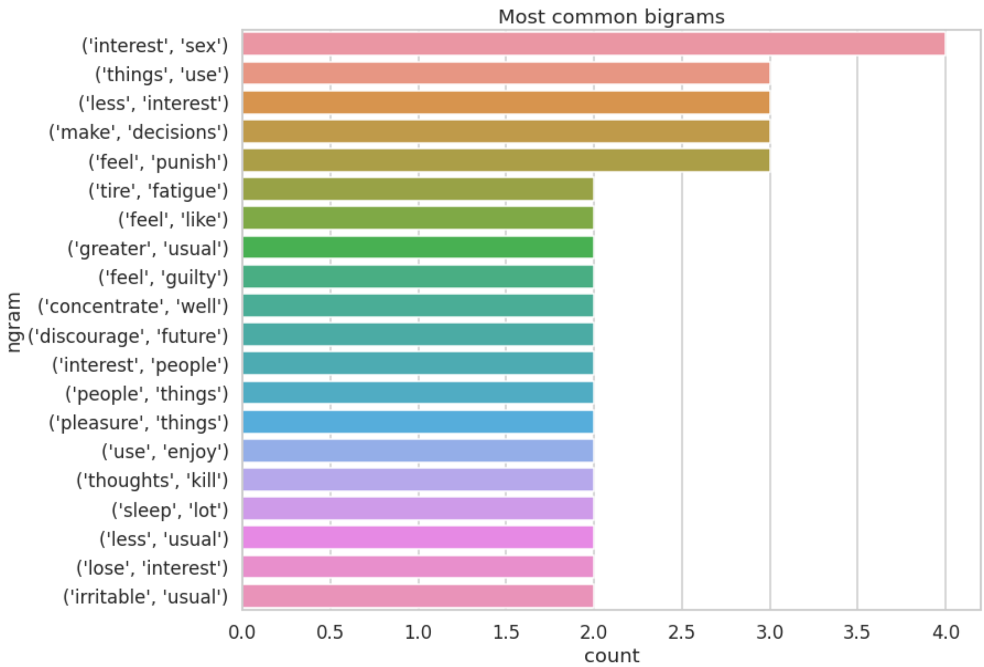

Figure 8: Most Common Bigram Tokens in BDI-II Questionnaire

4.2.2 Sentiment Analysis using sklearn model selection

In most of its questions, the original BDI-II questionnaire has a negative sentiment in the context of the questions asked to the patient, as 76.7% of the questions are negative while the rest are neutral. As part of the analysis, we have created a list of the most frequent tokens in the BDI-II questionnaire and how frequent they are used in each sentiment class as illustrated in the Figure 9.

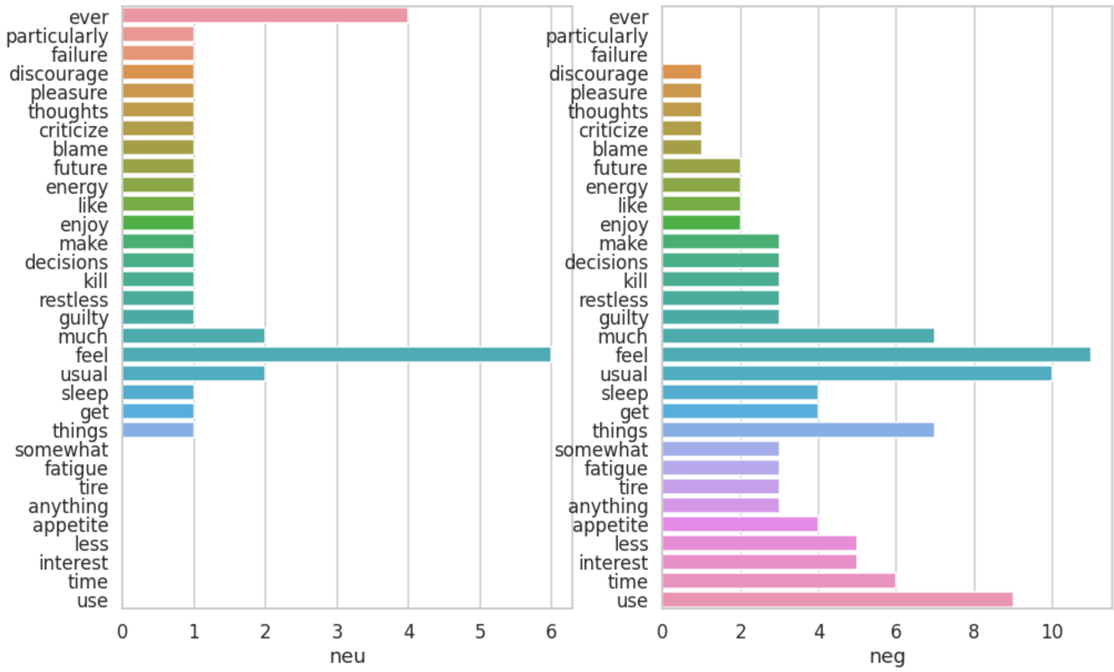

Figure 9: Most Frequent Tokens in BDI-II and Their Frequency in Each Sentiment Class

4.2.3 Co-occurrence Network

Co-occurrence network for the BDI-II questionnaire was created after stop words removal by creating a dictionary of token combinations. The dictionary consists of tuples of two tokens each and the frequency of the co-occurrence of this combination. The Itertools library is used to get the combinations of tokens in the questionnaire, and NetworkX library is used to represent the co-occurrence network as seen in Figure 10 where the network nodes are represented by the tokens in the dictionary tuples, and each of the tokens in a tuple are connected to each other in the network.

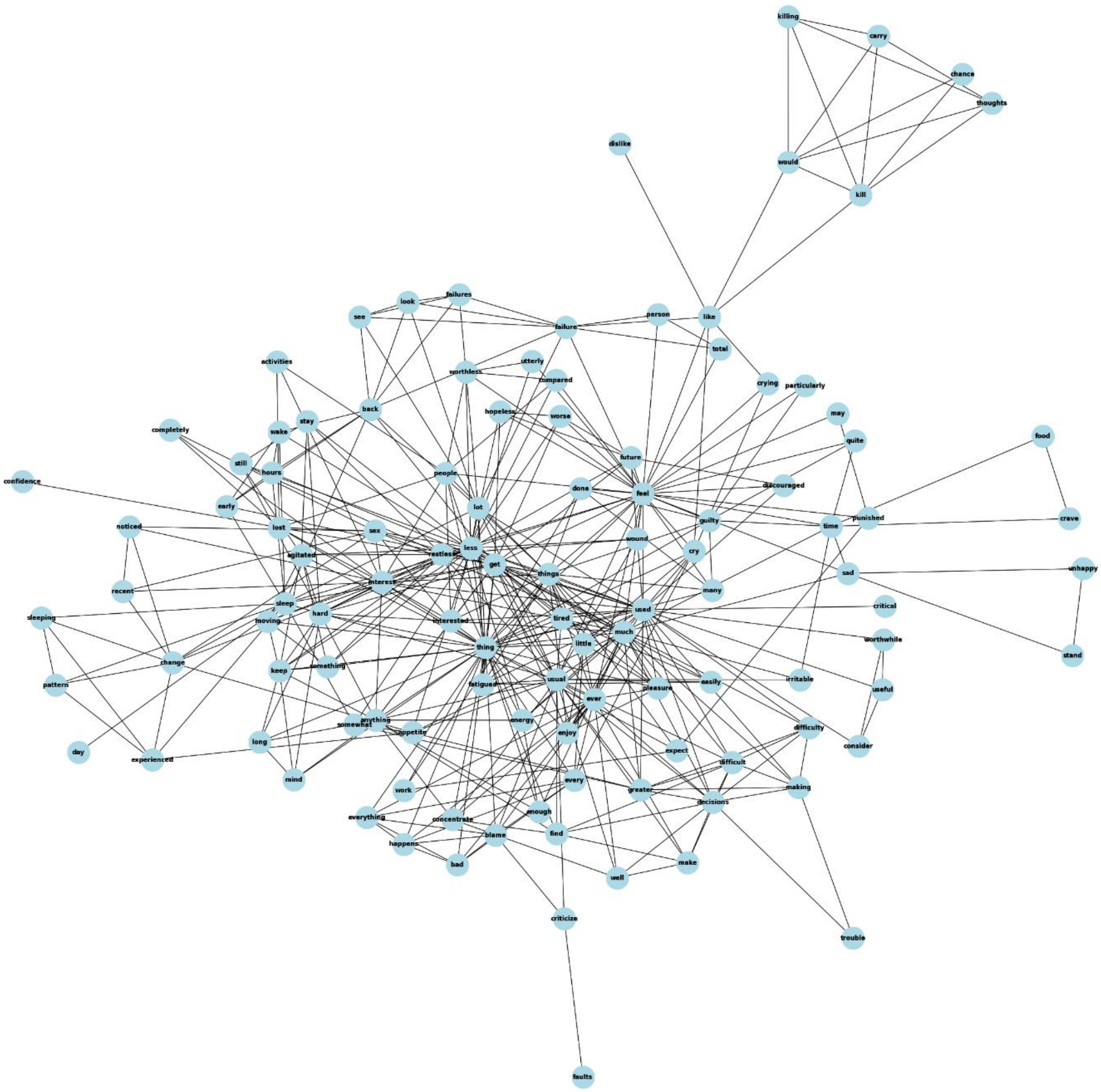

Figure 10:Co-occurrence of BDI-II Tokens

For easier visualization of the network, an interactive visualization of the network can also be shown as an html page in the web browser. Physics filter[43] allows more fluid graph interactions as seen in Figure 11.

[43] https://pyvis.readthedocs.io/en/latest/tutorial.html

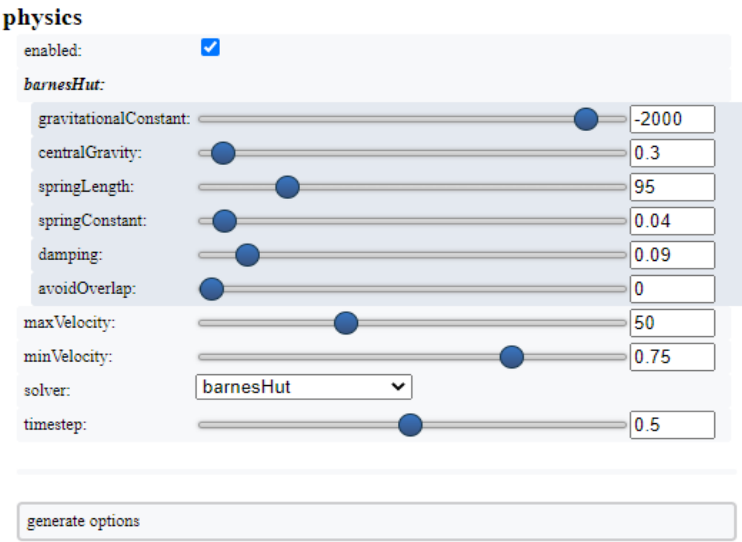

Figure 11: Physics Filter for html Visualization of NetworkX

4.3 Terms' Expansion Using Extrinsic Resources

In the next sections, we discuss the process of expanding the BDI-II questions with relevant medical terms that can be extracted from extrinsic medical knowledge resources, such as the UMLS. Our goal in this context is to evaluate the impact of expanding the original questions with relevant depression terms that can be further employed as part of the feature engineering phase. We argue that contextualizing and expanding the terms of the BDI-II questions does not only assist in producing more concise grouping of the questions, but also it can assist in identifying relevant symptoms and diseases to clinical depression.

4.3.1 Terms' Expansion Using UMLS Metathesaurus

Using "PyMedTermino" library in Python and UMLS APIs, a new questions dataset was created as seen in the partial dataset seen in Table 5. The original questionnaire consists of 21 questions with 4 answers each. The new questions dataset consists of 166 answers for the same 21 categories of questions. A lexical semantic network was created for these 166 answers based on the relationships in UMLS Metathesaurus.

Table 5: Partial UMLS Relationships for Output Dataset

Question	Relation	Related To
feeling sad in past 7d	has_class	Clinical
feeling sad in past 7d	has_component	Feeling sad
feeling sad in past 7d	has_supersystem	Patient
feeling sad in past 7d	has_time_aspect	7 days
feeling sad in past 7d	has_answer	I do not feel sad
feeling sad in past 7d	has_answer	I feel sad less than half the time
feeling sad in past 7d	has_answer	I feel sad nearly all of the time
I feel sad much of the time.	answer_to	Feel sad question
I am sad all the time.	answer_to	Feel sad question
I am so sad or unhappy that I can’t stand it.	answer_to	Feel sad question

Figures 12 and 13 respectively show the semantic relationships retrieved from UMLS metathesaurus database, and the top 20 relations for the new questions dataset.

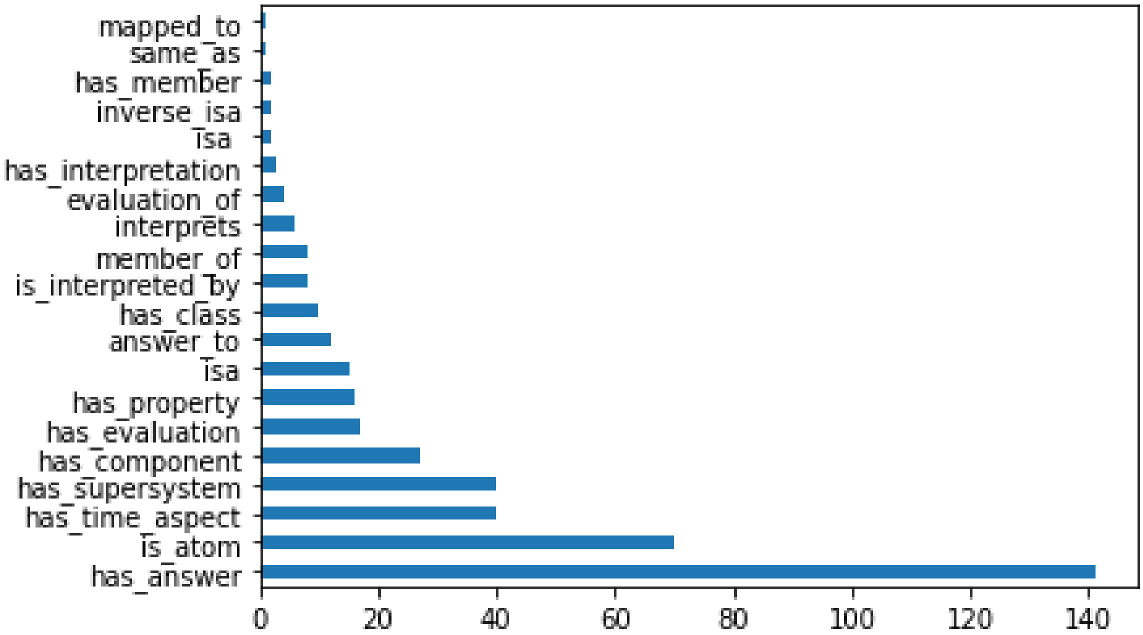

Figure 12: Bar Chart of UMLS Metathesaurus Semantic Relationships

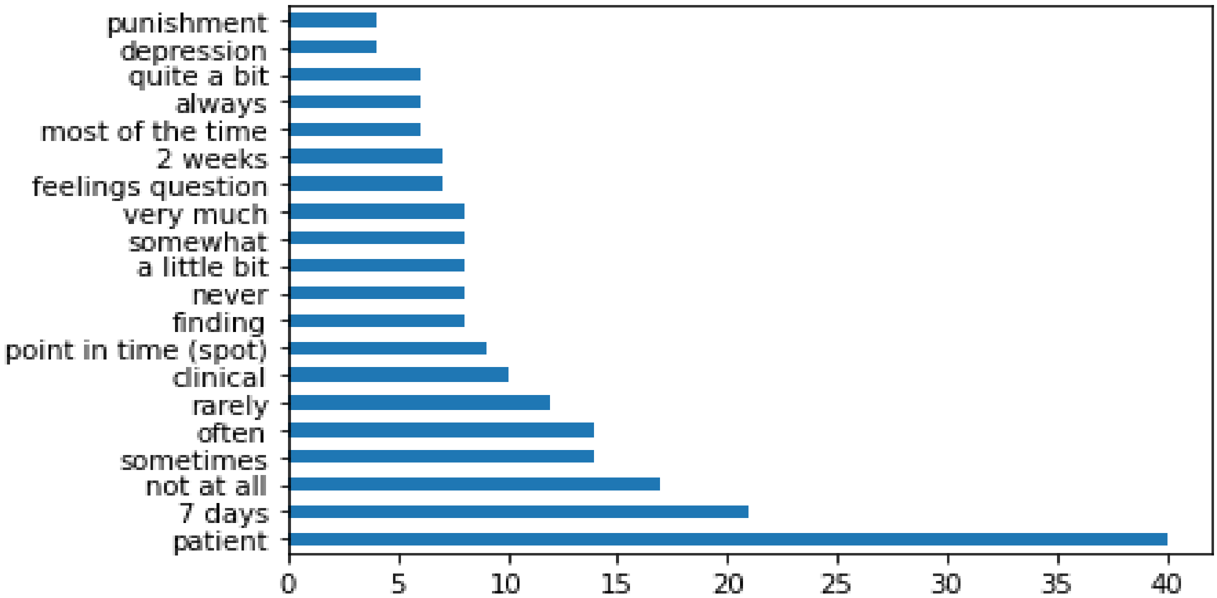

Figure 13: Bar Chart of Top 20 Answers to Questions Relationships

Relationships retrieved from UMLS Metathesaurus are defined in clinically used questionnaires like PROMIS (Patient Reported Outcomes Measurement Information

System), Rosenberg Self-Esteem Scale, SIQ (Suicidal Ideation Questionnaire), PHQ (Patient Health Questionnaire), CIDI-SF (The Composite International Diagnostic Interview), Neuro-QOL (Quality of Life), FACIT (Functional Assessment of Chronic Illness Therapy), and GDS (Geriartic Depression Scale).

Based on the above, a lexical semantic network was created. Figure 14 shows a partial view of the lexical semantic network for relationships retrieved from UMLS Metathesaurus database.

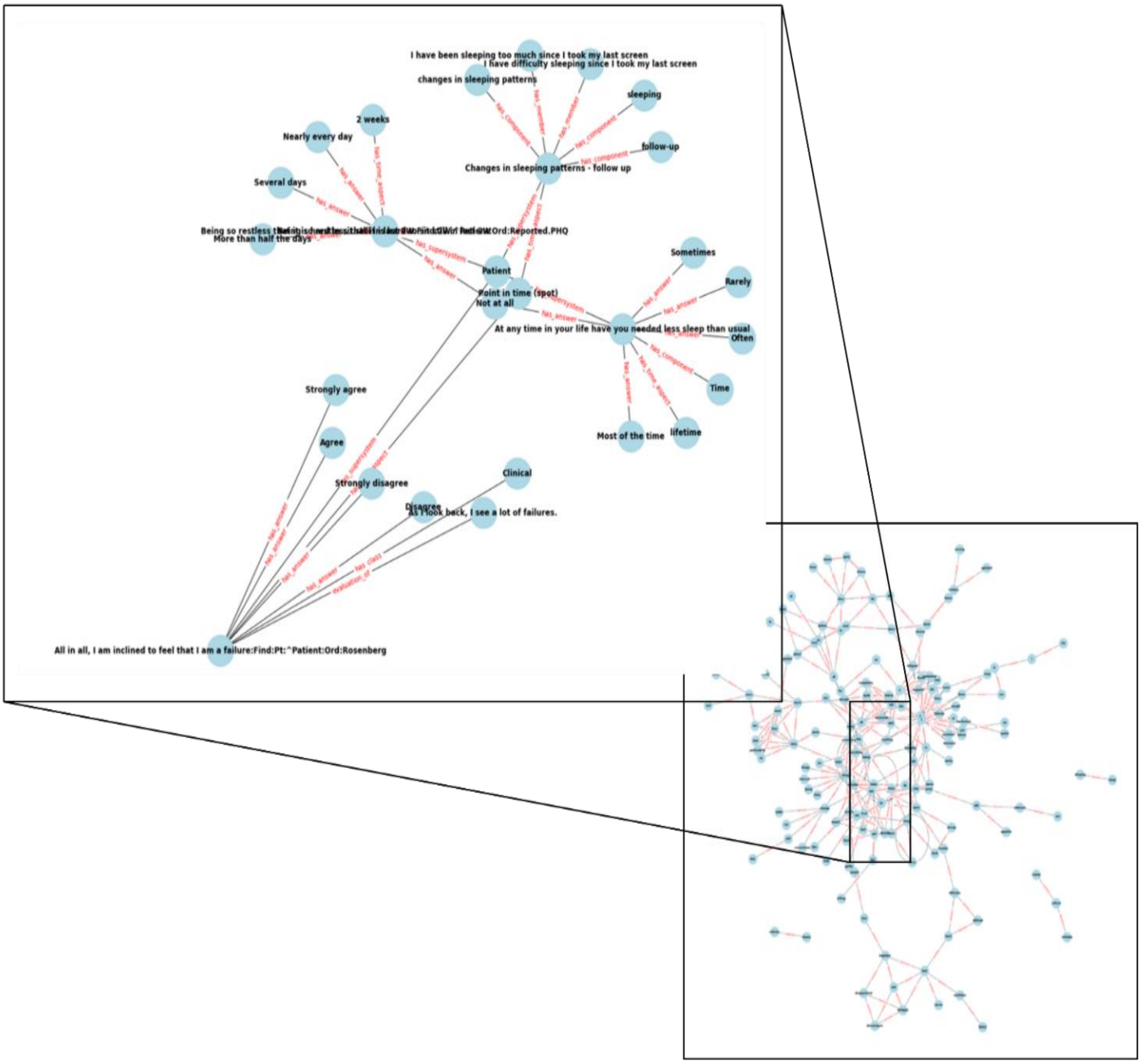

Figure 14: Partial Lexical Semantic Network Output

4.3.2 Terms' Expansion Using NLTK WordNet

Terms' expansion was performed using NLTK wordnet. For that purpose, the list of unique terms was used to get all the synonyms, hypernyms and meronyms using wordnet, which resulted in an expanded set of tokens. The figure below represents a word cloud of the most frequent tokens in the expanded list of terms. As seen in the word cloud, many tokens are not clinical text and are not related to the depression screening context. Therefore, further enrichment and enhancement is required using clinical text interpretation in Python.

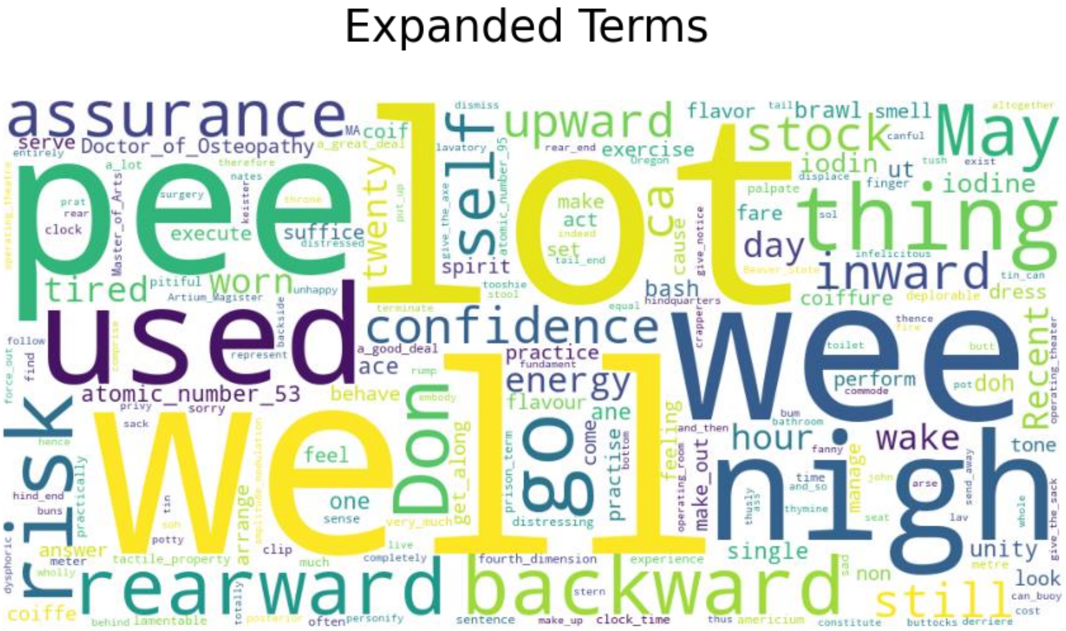

Figure 15: WordCloud Representing the Most Frequent Tokens in the Expanded Set Using WordNet

4.3.3 Terms' Expansion Using Stanza

Stanza is a collection of tools used in linguistic analysis of human languages.[44] Stanza library contains several biomedical models that are used to annotate clinical text. In our work we used "Mimic" and "i2b2" packages in "Stanza" to annotate clinical text in the BDI-II questionnaire and its interpretation in the model as seen in the code snapshot shown in Figure 16.

[44] https://stanfordnlp.github.io/stanza/

```
for question in questions:
  my_sentiment= nlp_biomedical_mimic_pipe(question)
  for ent in my_sentiment.entities:
    print(f'{ent.text} \t {ent.type}')
```

```
sad      PROBLEM
sad      PROBLEM
a failure        PROBLEM
failures         PROBLEM
a total failure          PROBLEM
crying   PROBLEM
more restless    PROBLEM
wound    PROBLEM
more restless    PROBLEM
wound    PROBLEM
restless         PROBLEM
agitated         PROBLEM
restless         PROBLEM
agitated         PROBLEM
more irritable   PROBLEM
irritable        PROBLEM
more irritable   PROBLEM
any change in my appetite        PROBLEM
1    I can’t concentrate         TREATMENT
fatigued         PROBLEM
fatigued         PROBLEM
too tired        PROBLEM
fatigued         PROBLEM
too tired        PROBLEM
fatigued         PROBLEM
```

Figure 16: Annotating Clinical Text in BDI-II Questionnaire

Tokens that are annotated as clinical text in BDI-II questionnaire are given higher weight than the rest of the text and are then used to enrich the BoW model.

4.3.4 Terms’ Expansion Using eRisk 2020 Dataset

Comparing unique terms in BDI-II with text from eRisk dataset (from 70 subjects for depression screening): Out of 173 unique terms used in BDI || questionnaire 171 terms were also used by the people who were subject of study in eRisk Dataset which contains 29918 unique terms. This indicates that terms used in the BDI-II are also used by people diagnosed with depression when expressing themselves on their social media accounts. A list of the most frequent terms used by the studied subjects in the eRisk dataset is created to be used to enrich the BoW model. The list contains the following tokens:

like, do, one, would, people, get, think, rule, time, comment, thing, know, also, make, really, see, even, want, removed, work, lot, year, good, much, way, have, that, well, still, could, go, first, many, something, need, day, use, world, you, say, post, instead, discussion, help, breaking, consider, daily, Hitler, never, love, friend, thanks, appeal, process, message, must, notice, clicking, moderator, within, page, inform

4.4 Bag of Words Model

After creating the expanded lexical semantic network using BDI II, WordNet and UMLS results; NLTK probability and FreqDist libraries are used to apply Binary Scoring for terms occurrences in BDI II questionnaire, UMLS expanded dataset, and the most common terms used by people diagnosed in clinical depression on their social media accounts -for that eRisk dataset is used-. Bag-Of-Words (BoW) is then created for the terms occurring in all three datasets.

For better results, Bag-Of-Words Model using Scikit-Learn and Count Vectorizer libraries was applied after stop words removal:

ability	able	activity	afraid	agitated	agree	almost	also	always
0	0	1	0	1	0	0	0	0
1	1	1	1	1	1	1	0	1
0	0	0	0	0	0	0	1	0

Figure 17: Bag of Words Using SciKit-Learn

Using the BoW model, a list of the most frequent terms is created. This list of tokens and their semantically related tokens are used to reformulate the questions of BDI II questionnaire and weighting them based on their usage and frequency in other contexts.

4.5 Questions' Re-weighting

The most frequent tokens that are used in the BDI-II and the extrinsic resources used to enhance it based on the BoW model are used to reformulate the English version of BDI-II questionnaire. The output of the BoW model resulted in 20 tokens that are highly used in depression screening questionnaires and by people diagnosed with depression on their social media accounts. These tokens and their semantically related tokens are given higher weight in the questions' order and grouping. These terms are:

change, consider, cry, decision, get, like, lot, make, many, much, pattern, people, see, something, still, thought, time, well, work, and would.

4.6 Questions' Clustering and Re-grouping

After creating a semantically enhanced formulation of depression screening questions. These questions are presented to a psychiatrist to enhance the formulation and decide on whether a certain question can or cannot be asked to a patient in the formulation that is created using our lexical semantic network.

All 21 categories were used in mapping the questions into 6 clusters using the classification models. It is noticed the expanded dataset isn't balanced as seen in Figure 18.

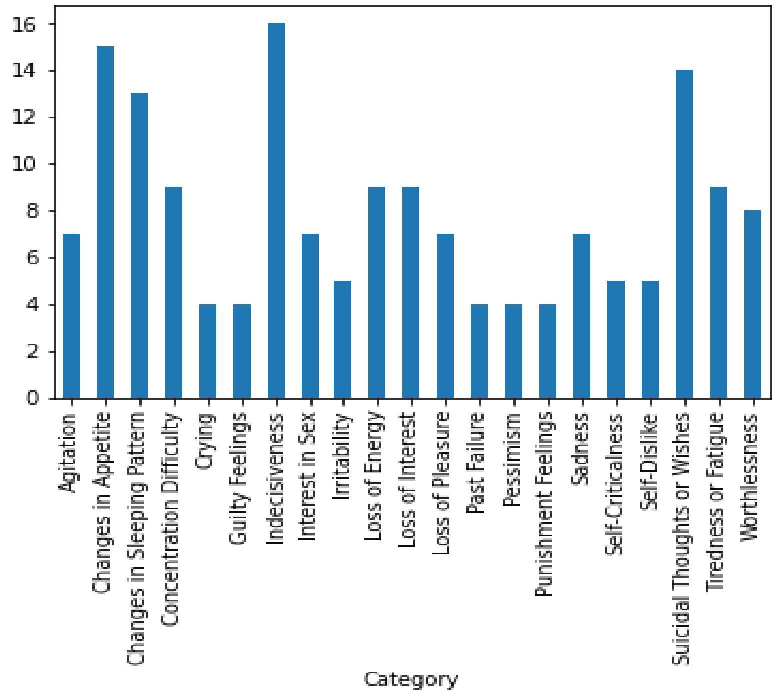

Figure 18: Expanded Questions Frequency Per Category

To create a new classification for the expanded questions; multiple classification models were applied to reformulate the order of questions appearing in the questionnaire based on the category they belong to. For that K-means and TF-IDF models were applied, and questions were mapped into 21 clusters as in the original BDI II questionnaire as seen in Figure 19.

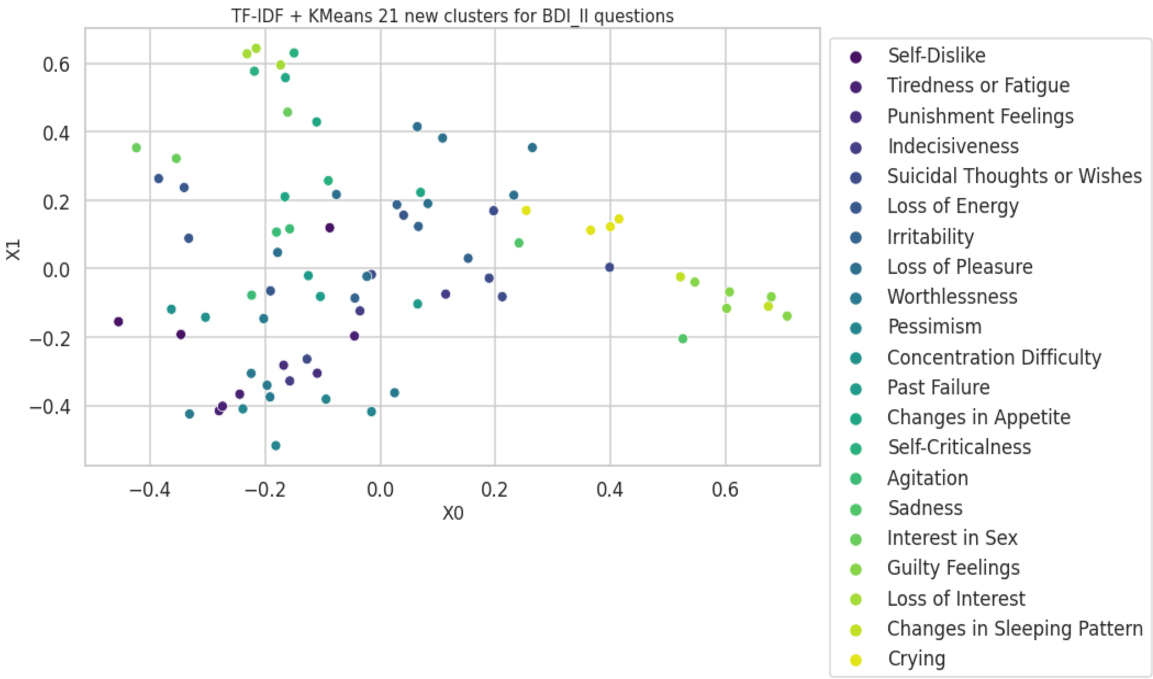

Figure 19: Text Clustering Using TF-IDF and K-Means for BDI-II Questions

Similarly, the expanded set of questions from UMLS dataset are also clustered into 21 clusters as in the original BDI-II questionnaire as seen in Figure 20.

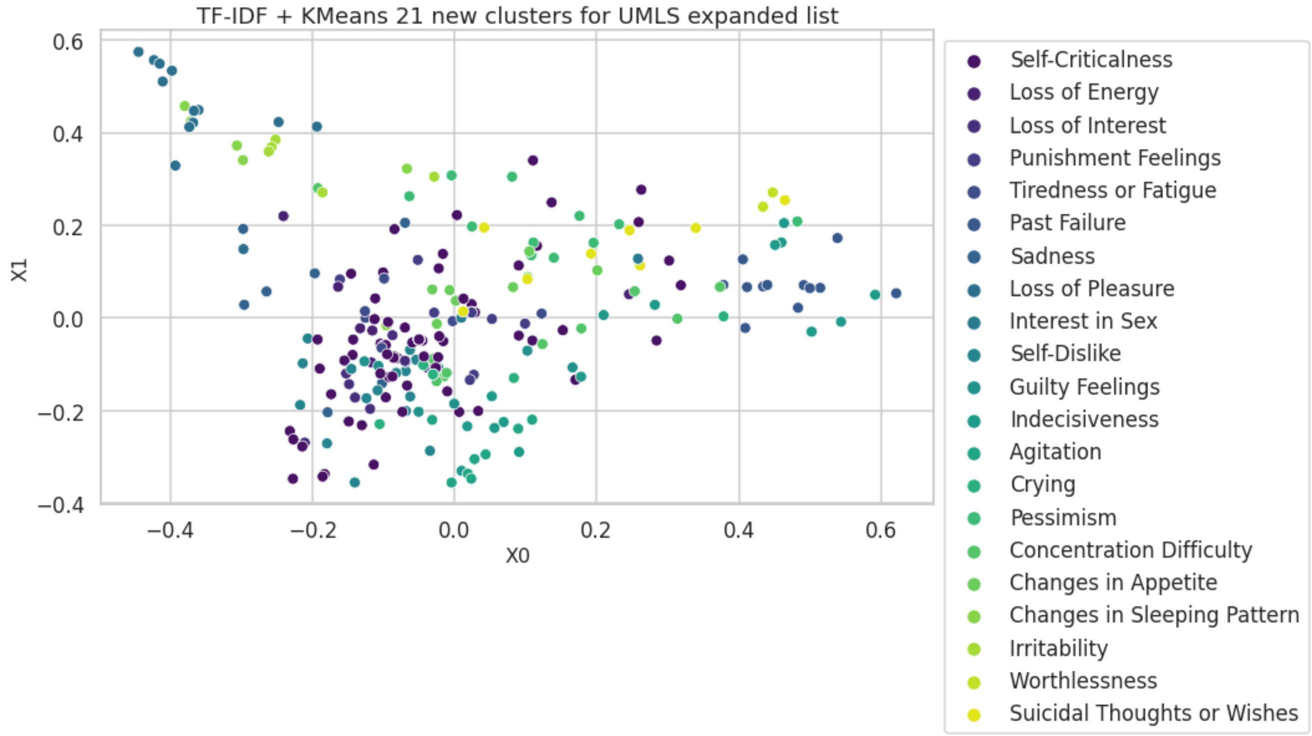

Figure 20: Text Clustering Using TF-IDF and K-Means for UMLS Expansion List of Questions

4.7 BDI Arabic Version Analysis

Using camel-tools library[45], the Arabic version of the BDI questionnaire was pre-processed in a similar pipeline to the one used to pre-process the English version of BDI-II. Arabic data pre-processing includes dediacritization, text normalization, tokenization, and punctuation removal as explained below:

1. Arabic BDI questionnaire was inputted into Python notebook in the form of a csv file that was then imported into a DataFrame using pandas[46] Python library.
2. Dediacritization[47] was applied to all questions by removing all short vowel marks (the harakat/ motions) from the Arabic text.

[45] https://camel-tools.readthedocs.io/en/latest/api/utils.html
[46] https://pandas.pydata.org/

Before Dediacritization
إنني لست متشائماً بشأن المستقبل أشعر بأنني أستحق العقاب أحياناً
After Dediacritization
إنني لست متشائما بشأن المستقبل أشعر بأنني أستحق العقاب أحيانا

3. Text normalization[48] was applied to all questions by removing all glottal stop signs (hamza) from the text as they are treated as special characters in the Arabic language. In this step all variations of the Arabic letter (Alef) were normalized to plain alef using normalize_alef_ar tool in camel_tools Python library. In addition to that, all occurrences of the Arabic letter (teh marbuta) were normalized to a (heh) character, and all occurrences of (Alef Maksura) were normalized to (yeh) using normalize_alef_maksura_ar and normalize_teh_mabuta_ar tools.

Before Text Normalization
ما زالت الأشياء تعطيني شعورا بالرضى كما كانت عادة
After Text Normalization
ما زالت الاشياء تعطيني شعورا بالرضي كما كانت عاده

4. Special characters, numbers and white spaces were removed.
5. Clean data was then tokenized using simple_word_tokenize[49] tool in camel_tools Python Library resulting into 691 Arabic tokens, out of which there are 304 unique tokens including:

[47] https://camel-tools.readthedocs.io/en/latest/api/utils/dediac.html
[48] https://camel-tools.readthedocs.io/en/latest/api/utils/normalize.html
[49] https://camel-tools.readthedocs.io/en/latest/api/tokenizers/word.html

اشعر، بالحزن، الاوقات، طيله، الوقت، استطيع، التخلص،الشعور، التعاسه، لدرجه، اتحملها، لست، متشائما، بشان، المستقبل، بالتشاؤم، يوجد، لدي، اطمح، للوصول، اليه، امل، فيه، وهذا، الوضع، الممكن، تغييره، بانني، فاشل، بشكل، عام، اواجه، الفشل، اكثر، مما، يواجه، الانسان، لعادي، عندما، انظر، الي، حياتي، الماضيه، اراه، الكثير، انسان، فشلا، تاما، زالت، الاشياء، تعطيني، شعورا، بالرضي، كما، كانت، عاده، بمتعه، علي، النحو، الذي، كنت، اعد، بايه، متعه، حقيقيه، اي، شيء، الاطلاق، شعور، بعدم، الرضا، والملل، بالذنب، معظم، كافه، استحق، عقابا، نوع، العقاب، احيانا، كثيرا، احس، اعاقب، واعذب، وانني، ذلك، بخيبه، الامل، نفسي، اكره، اسوا، الاخرين، انتقد، بسبب، اخطاء، وضعف، الوم، طيله، اخطائي، يحدث،علاقه، مباشره، بذلك، لايوجد، ايه، افكار، انتحاريه، توجد، الافكار، الانتحاريه، ولكني، اقوم، بتنفيذها، ارغب، قتل، ساقتل، توفرت، الفرصه، السانحه، لذلك، ابكي، المعتاد، هذه، الايام، لقد، قدره، البكاء، ولكنني، استطيع، بالرغم، الصحية، المعدة

4.7.1 Arabic Tokens Translation Using Translators Library

To expand the Arabic set of terms, all unique terms are translated to English using "translators[50]" library in Python. Google translate[51] was applied using language_map function in "translators" library. The set of translated tokens of BDI are then compared to the original set of tokens of BDI-II. New terms are added to the expanded English list of terms based on the translation of the Arabic terms, as in the follow examples:

Arabic Term	Translation
الحزن	Sorrow
التخلص	Get rid of
التعاسة	Misery
أطمح	I aspire
الرضى	Satisfaction
أعذب	Tormented
ضعف	Weakening

4.7.2 Arabic Tokens Expansion

After defining the expansion terms that are used in the reformulation of the English version of BDI-II, the Arabic version of BDI questionnaire will also be enhanced using

[50] https://pypi.org/project/translators/
[51] https://translate.google.com/

the corresponding Arabic terms as in the English version based on the same extrinsic resources used to enrich the English version. For that, the English expansion tokens are translated back to Arabic to create the Arabic list of expanded tokens that are then used to enrich the Arabic questionnaire.

4.8 Results and Discussion

In this chapter, we presented the expansion methods used to enrich the BDI-II questionnaire and discover relationships among terms of the questions using UMLS, Wordnet and e-Risk. The original and expanded relationship terms formulated a knowledge graph that provides a shared concept categorization for each cluster. In our approach, we kept the number of questions as in the original BDI-II questionnaire. However, the questions are reformulated using the terms with higher weights based on their occurrences in multiple extrinsic resources.

Using the most common terms in each of the extrinsic resources along with the original terms of BDI-II, we have reformulated the questions as seen in Table 6.

Table 6: BDI-II Questions' Reformulation Example

Question before reformulation	**Question after reformulation**
Sadness	
I do not ~~feel~~ sad.	**I am not** sad.
I ~~feel~~ sad much of the time.	I am sad much of the time.
I am sad all the time.	I am sad all the time.
I am so sad or unhappy that I cannot stand it.	I am so sad or unhappy that I cannot stand it.
Pessimism	
I am **not discouraged** <u>about my</u> future	I have **not given up hope** <u>for the</u> future.
I feel **more discouraged** about <u>my future</u> than **I used to be.**	I am **more pessimistic** about <u>the future</u> than **before.**

I do not **expect** things to work out for me.	I do not **have high hopes** for things to work out for me.
~~I feel~~ my future is hopeless and ~~will~~ only get worse.	My future is hopeless and only gets worse.
Past failure	
I do not ~~feel~~ like a failure.	I am not a failure.
I have failed **more than I should have.**	I have failed *much.*
As I look back, I see a lot of failures.	**I made lots of mistakes in the past.**
I ~~feel~~ I am a total failure as a person.	As a person, I am a total failure.

As observed from the example above, the word "feel" was removed from the questions as it showed a lower weight among the rest of the tokens while words like "much" and "time" were given a higher weight in the questions and thus were not removed from the questions. As a result of the reformulation some questions were not changed like "I am sad all the time", while others were partially or completely reformulated like "I have failed more than I should have" and "As I look back, I see a lot of failures".

BDI-II questions are also being analyzed to be re-grouped into new clusters. As multiple clustering techniques were applied, we got different clusters for each method. In K-Means clustering, the model scored an accuracy of 0.61. Each question was recategorized under a category with keeping the number of original categories of BDI-II resulting into a change in the number of questions under each category as seen in the table below:

Table 7: K-Means Clustering for BDI-II Questions

Group	Original Topic	Original number of questions	Resulting number of questions
1	Sadness	4	3
2	Pessimism	4	3
3	Past Failure	4	3
4	Loss of Pleasure	4	4
5	Guilty Feelings	4	3
6	Punishment Feelings	4	5
7	Self-Dislike	4	2
8	Self-Criticalness	4	3
9	Suicidal Thoughts or Wishes	4	4
10	Crying	4	4
11	Agitation	4	1
12	Loss of interest	4	5
13	Indecisiveness	4	7
14	Worthlessness	4	4
15	Loss of Energy	4	3
16	Changes in Sleeping Pattern	7	6
17	Irritability	4	7
18	Changes in Appetite	7	10
19	Concentration Difficulty	4	4
20	Tiredness or Fatigue	4	5
21	Interest in Sex	4	4

Table 8 provides an example of the new categorization of the questions using K-Means:

Table 8: Example of BDI-II Questions' New Categories Using K-Means Clustering

BDI-II Question	Original Category	K-Means Clustering Category
I do not feel sad	Sadness	Sadness
I am sad all the time	Sadness	Sadness
I am sad all the time.	Sadness	Irritability
I am so sad or unhappy that I can’t stand it	Sadness	Sadness
I am not discouraged about my future	Pessimism	Pessimism
I feel more discouraged about my future than I used to be	Pessimism	Pessimism
I do not expect things to work out for me	Pessimism	Pessimism
I feel my future is hopeless and will only get worse	Pessimism	Changes in Appetite
I do not feel like a failure	Past Failure	Worthlessness
I have failed more than I should have	Past Failure	Past Failure
As I look back, I see a lot of failures	Past Failure	Past Failure
I feel I am a total failure as a person	Past Failure	Past Failure

TF-IDF has scored better accuracy than K-Means as the accuracy of TF-IDF on BDI-II questions is 0.72. The questions were regrouped into different categories with a resulting change in number of questions under each category as seen in Table 9.

Table 9: TF-IDF Clustering for BDI-II Questions

Group	Topic	Original number of questions	Resulting number of questions
1	Sadness	4	2
2	Pessimism	4	4
3	Past Failure	4	3

4	Loss of Pleasure	4	8
5	Guilty Feelings	4	7
6	Punishment Feelings	4	3
7	Self-Dislike	4	6
8	Self-Criticalness	4	3
9	Suicidal Thoughts or Wishes	4	6
10	Crying	4	4
11	Agitation	4	4
12	Loss of interest	4	3
13	Indecisiveness	4	3
14	Worthlessness	4	7
15	Loss of Energy	4	6
16	Changes in Sleeping Pattern	7	4
17	Irritability	4	4
18	Changes in Appetite	7	4
19	Concentration Difficulty	4	2
20	Tiredness or Fatigue	4	4
21	Interest in Sex	4	3

The BDI-II questionnaire was also clustered per questions' categories as many of these categories may share some common characteristics, and therefore be grouped under a larger group. For that TF-IDF was used to create 3 new groups for the BDI-II categories and the results were as seen in the Table 10.

Table 10: TF-IDF Clustering for BDI-II Categories

Cluster number	Categories	Number of BDI-II categories in cluster	Keywords
Cluster 1	Sadness Pessimism Past Failure Loss of Pleasure Guilty Feelings Punishment Feelings Crying Worthlessness Loss of Energy Irritability Concentration Difficulty	11	People, enjoy, guilty, thing, time
Cluster 2	Self-Dislike Self-Criticalness Suicidal Thoughts or Wishes Indecisiveness	4	Blame, kill, myself, decision
Cluster 3	Agitation Loss of interest Changes in Sleeping Pattern Tiredness or Fatigue Changes in Appetite Interest in Sex	6	People, sleep, appetite, restless.

The results that we got show that the first group contains almost all the questions that are related to feelings (e.g., sadness and pessimism). The second group contains questions that are related to self-harm -physically and psychologically- (e.g., self-dislike and suicidal thought and wishes) as these questions are mostly related to thoughts about the self. And the third group contains questions that are related to general habits and activities that the patient is used to (e.g., loss of interest in activities and changes in sleeping pattern).

4.9 Summary

In this chapter, we presented how we applied each of the enrichment approaches on the BDI-II questionnaire. The first enrichment approach was using UMLS metathesaurus to expand the terms based on medical metathesaurus showing the relationships between these questions and others in the metathesaurus. The second enrichment approach was using Wordnet to expand the list of terms based on lexical semantic relations as seen in Wordnet. The third enrichment approach was using eRisk dataset to analyze social media posts that are posted by people that already have screening results for depression. A list of words was then identified by applying the Bag-Of-Words model on BDI-II original tokens along with the expanded list of tokens that resulted from all three enrichment approaches.

After identifying the set of tokens that resulted from the Bag-Of-Words model, the most common words on each of the approaches was identified and the tokens that occurred in all datasets were given higher weight and were used later to reformulate the original questions. The reformulation was manually done after identifying the key words that must be used in the questionnaire. The number of questions was not changed during the reformulation process, only the context of each question was enhanced.

In the final stage of our project, pre-trained clustering models were used to re-group the questions into new categories, and therefore reformulate the original questionnaire's categories accordingly. TF-IDF provided better results when applied on categories rather than on questions.

The reformulated questions of BDI-II questionnaire and the new categorization were presented to a psychiatrist who evaluated the efficiency of the syntax of the reformulated questions and the newly used terminologies. As part of this evaluation, the new version of the questionnaire will be presented to other psychiatrists to evaluate it towards its usage in the clinical depression screening process.

Chapter 5 - Conclusions and Future Work

This chapter outlines our strategy, explains our findings, and highlights the contributions we made by using the proposed tool. Furthermore, it emphasizes the tool's prospective enhancements for the existing implementation. The chapter is organized into two sections: 5.1, which covers our research work and highlights the methodologies and approaches employed in the proposed system, and 5.2, which covers future work and issues that we intend to address in future system updates.

5.1 Conclusions

To screen depression, psychologists recommend filling out a questionnaire with a symptom rating scale. One of the widely used depression screening questionnaires is the BDI-II questionnaire. The questionnaire which was last changed in 1996 has a large number of categories that might not all be convenient for the patient who is taking the test. The strength of our project is that it was the first attempt of applying lexical semantic based clustering on BDI-II questions. We aimed to explore how employing knowledge acquired by medical resources might improve the result in this situation.

We proposed BDI-II Questions reformulation and clustering techniques, for this, we used pre-trained clustering techniques to obtain new patterns and cluster the BDI-II questions under semantically enhanced categories. For that, we have constructed a knowledge graph employing synonymy/ meronymy relationships to link tokens derived from original BDI-II questions and the lexical and medical knowledge resources that were used in this project.

First, the output lexical semantic network represents the tokenized group of BDI-II questions and the relationships between these tokens and the list of extrinsically expanded tokens in the network. Second and after identifying the relationships between the tokens, a list of terms was extracted using the Bag-Of-Words model to represent the keywords that were used to reformulate the questions by giving these terms a higher weigh and importance in the questionnaire. Third, we have utilized two different clustering models for re-grouping the original questions of BDI-II questionnaire. Without using any knowledge resources, K-Means and TF-IDF were utilized to cluster the questions by grouping similar questions into distinct categories. The goal was to demonstrate how medical knowledge can aid in improving the quality of clustering methods.

The newly proposed categories, which are semantically enhanced and based on medical knowledge graphs, have produced a new categorization of questions that are close in their semantic distance and have less overlap that the original BDI-II questions that are being used by psychologists and healthcare professionals to screen depression.

5.2 Challenges and Future Work

Although the experiment yielded encouraging results, there are some improvements that can be made to the methodologies given in this research. These can be summarized as follows:

- Improve clustering techniques based on pre-trained language models like BERT/BIOBERT to include broader medical enrichment relationships during the processing.

- Explore additional relationships that can be retrieved from UMLS or other medical resources so that we can expand our depression screening terms, and then discover lexical patterns between the questions set.
- Evaluate our approach with a larger dataset and ask psychologists to test and evaluate the outcome.
- Link the symptoms that are screened as part of the depression screening process to other diseases that can be diagnosed by these symptoms.
- Build an automated depression screening model based on the new categorization.
- Investigate with healthcare professionals the possibility of enriching the questionnaire's terminology by adding a free-text question where the patient expresses his/her feelings. The terms that are written by the patient are then used as an input to enrich depression screening questions.

References

Alansari, B. (2005). "Beck Depression Inventory (BDIII) sourcebook of personality disorders scales." Kuwait: The New Book Home Co., Kuwait University.

Arnaez, J. M., et al. (2020). "The association of depression stigma with barriers to seeking mental health care: A cross-sectional analysis." Journal of Mental Health 29(2): 182-190.

Beck, A. T., et al. (1996). "Comparison of Beck Depression Inventories-IA and-II in psychiatric outpatients." Journal of personality assessment 67(3): 588-597.

Brown, C., et al. (2010). "Depression stigma, race, and treatment seeking behavior and attitudes." Journal of community psychology 38(3): 350-368.

Buzydlowski, J. W. (2015). "Co-occurrence analysis as a framework for data mining." Journal of Technology Research 6: 1-19.

Calvo, R. A., et al. (2017). "Natural language processing in mental health applications using non-clinical texts." Natural Language Engineering 23(5): 649-685.

Chen, Z., et al. (2018). "Evaluating semantic relations in neural word embeddings with biomedical and general domain knowledge bases." BMC medical informatics and decision making 18(2): 53-68.

Chong, W. Y., et al. (2014). Natural language processing for sentiment analysis: an exploratory analysis on tweets. 2014 4th international conference on artificial intelligence with applications in engineering and technology, IEEE.

Clement, S., et al. (2015). "What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies." Psychological medicine 45(1): 11-27.

Colligan, E. M., et al. (2020). "Barriers and facilitators to depression screening in older adults: a qualitative study." Aging & mental health 24(2): 341-348.

Corrigan, P. W., et al. (2014). "The impact of mental illness stigma on seeking and participating in mental health care." Psychological Science in the Public Interest 15(2): 37-70.

Cronly, J., et al. (2018). "Online versus paper-based screening for depression and anxiety in adults with cystic fibrosis in Ireland: a cross-sectional exploratory study." BMJ open 8(1): e019305.

Dancygier, B. (2017). The Cambridge handbook of cognitive linguistics, Cambridge University Press.

Dockery, L., et al. (2015). "Stigma-and non-stigma-related treatment barriers to mental healthcare reported by service users and caregivers." Psychiatry Research 228(3): 612-619.

Eack, S. M., et al. (2006). "Limitations of the Patient Health Questionnaire in identifying anxiety and depression in community mental health: many cases are undetected." Research on social work practice 16(6): 625-631.

Elbedour, S., et al. (2007). "Post-traumatic stress disorder, depression, and anxiety among Gaza Strip adolescents in the wake of the second Uprising (Intifada)." Child abuse & neglect 31(7): 719-729.

Farruque, N., et al. (2019). Augmenting semantic representation of depressive language: From forums to microblogs. Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer.

García-Batista, Z. E., et al. (2018). "Validity and reliability of the Beck Depression Inventory (BDI-II) in general and hospital population of Dominican Republic." PloS one 13(6): e0199750.

Ghareeb, A. (2000). "Manual of the Arabic BDI-II." Alongo Press. Cairo Inventory: the author's twenty-five years of evaluation. Clin Psychol Rev 8: 77-100.

Guillot-Valdés, M., et al. (2019). "Reliability and validity of the Basic Depression Questionnaire." International Journal of Clinical and Health Psychology 19(3): 243-250.

Gurung, D., et al. (2022). "Stigma against mental health disorders in Nepal conceptualised with a 'what matters most'framework: a scoping review." Epidemiology and Psychiatric Sciences 31.

Kaltenboeck, A. and C. Harmer (2018). "The neuroscience of depressive disorders: A brief review of the past and some considerations about the future." Brain and Neuroscience Advances 2: 2398212818799269.

Kanter, J. W., et al. (2008). "The nature of clinical depression: Symptoms, syndromes, and behavior analysis." The Behavior Analyst 31(1): 1-21.

Kavasis, I., et al. (2005). "Translation and validation of the Depression Outcomes Module (DOM) in Greece." International Journal of Methods in Psychiatric Research 14(4): 221-229.

Ko, J. Y., et al. (2013). "Barriers in the diagnosis and treatment of depression in women in the USA: where are we now?" Neuropsychiatry 3(1): 1.

Kozlowski, M. and H. Rybinski (2019). "Clustering of semantically enriched short texts." Journal of Intelligent Information Systems 53(1): 69-92.

Levis, B., et al. (2019). "Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis." bmj 365.

Luo, X. and S. Shah (2018). Concept embedding-based weighting scheme for biomedical text clustering and visualization. Applied Informatics, Springer

Madianos, M. G., et al. (2011). "Posttraumatic stress disorders comorbid with major depression in West Bank, Palestine: a general population cross sectional study." The European Journal of Psychiatry 25(1): 19-31.

Madianos, M. G., et al. (2012). "Major depression across West Bank: a cross-sectional general population study." International journal of social psychiatry 58(3): 315-322.

Maurer, D. M., et al. (2018). "Depression: screening and diagnosis." American family physician 98(8): 508-515.

Miller, P., et al. (2021). "The performance and accuracy of depression screening tools capable of self-administration in primary care: A systematic review and meta-analysis." The European Journal of Psychiatry 35(1): 1-18.

Mojtabai, R., et al. (2011). "Barriers to mental health treatment: results from the National Comorbidity Survey Replication." Psychological medicine 41(8): 1751-1761.

Morrey, C. P., et al. (2012). "A chemical specialty semantic network for the Unified Medical Language System." Journal of cheminformatics 4(1): 1-11.

Rahat, M., et al. (2012). "Psychometric properties of the Arabic version of the obsessive compulsive beliefs questionnaire-44 in a student population." Iranian journal of psychiatry 7(4): 184.

Rajput, A. (2020). Natural language processing, sentiment analysis, and clinical analytics. Innovation in Health Informatics, Elsevier: 79-97.

Rangrej, A., et al. (2011). Comparative study of clustering techniques for short text documents. Proceedings of the 20th international conference companion on World wide web.

Reátegui, R. and S. Ratté (2018). "Comparison of MetaMap and cTAKES for entity extraction in clinical notes." BMC medical informatics and decision making 18(3): 13-19.

Renganathan, V. (2017). "Text mining in biomedical domain with emphasis on document clustering." Healthcare informatics research 23(3): 141-146.

Sandhiya, R. and M. Sundarambal (2019). "Clustering of biomedical documents using ontology-based TF-IGM enriched semantic smoothing model for telemedicine applications." Cluster Computing 22(2): 3213-3230.

Savoy, M. L. and D. T. O'Gurek (2016). "Screening your adult patients for depression." Family Practice Management 23(2): 16-20

Schaeffer, A. M. and D. Jolles (2019). "Not missing the opportunity: Improving depression screening and follow-up in a multicultural community." The Joint Commission Journal on Quality and Patient Safety 45(1): 31-39.

Sharma, G., et al. (2021). "DepHNN: a novel hybrid neural network for electroencephalogram (EEG)-based screening of depression." Biomedical Signal Processing and Control 66: 102393.

Sheehan, A. M. and H. McGee (2013). "Screening for depression in medical research: ethical challenges and recommendations." BMC Medical Ethics 14(1): 1-4.

Subakti, A., et al. (2022). "The performance of BERT as data representation of text clustering." Journal of big Data 9(1): 1-21.

Vogel, D. L., et al. (2011). "“Boys don't cry”: Examination of the links between endorsement of masculine norms, self-stigma, and help-seeking attitudes for men from diverse backgrounds." Journal of counseling psychology 58(3): 368.

Appendix 1: BDI-II questionnaire

1. BDI-II questionnaire

This questionnaire consists of 21 groups of statements. Please read each group of statements carefully, and then pick out the one statement in each group that best describes the way you feel. If several statements in the group seem to apply equally well, choose the highest number for that group.

1. Sadness

0. I do not feel sad.

1. I feel sad much of the time.

2. I am sad all the time.

3. I am so sad or unhappy that I can't stand it.

2. Pessimism

0. I am not discouraged about my future.

1. I feel more discouraged about my future than I used to be.

2. I do not expect things to work out for me.

3. I feel my future is hopeless and will only get worse.

3. Past Failure

0. I do not feel like a failure.

1. I have failed more than I should have.

2. As I look back, I see a lot of failures.

3. I feel I am a total failure as a person.

4. Loss of Pleasure

0. I get as much pleasure as I ever did from the things I enjoy.

1. I don't enjoy things as much as I used to.

2. I get very little pleasure from the things I used to enjoy.

3. I can't get any pleasure from the things I used to enjoy.

5. Guilty Feelings

0. I don't feel particularly guilty.

1. I feel guilty over many things I have done or should have done.

2. I feel quite guilty most of the time.

3. I feel guilty all of the time.

6. Punishment Feelings

0. I don't feel I am being punished.

1. I feel I may be punished.

2. I expect to be punished.

3. I feel I am being punished.

7. Self-Dislike

0. I feel the same about myself as ever.

1. I have lost confidence in myself.

2. I am disappointed in myself.

3. I dislike myself.

8. Self-Criticalness

0. I don't criticize or blame myself more than usual.

1. I am more critical of myself than I used to be.

2. I criticize myself for all of my faults.

3. I blame myself for everything bad that happens.

9. Suicidal Thoughts or Wishes

0. I don't have any thoughts of killing myself.

1. I have thoughts of killing myself, but I would not carry them out.

2. I would like to kill myself.

3. I would kill myself if I had the chance.

10. Crying

0. I don't cry any more than I used to.

1. I cry more than I used to.

2. I cry over every little thing.

3. I feel like crying, but I can't.

11. Agitation

0. I am no more restless or wound up than usual.

1. I feel more restless or wound up than usual.

2. I am so restless or agitated that it's hard to stay still.

3. I am so restless or agitated that I have to keep moving or doing something.

12. Loss of Interest

0. I have not lost interest in other people or activities.

1. I am less interested in other people or things than before.

2. I have lost most of my interest in other people or things.

3. It's hard to get interested in anything.

13. Indecisiveness

0. I make decisions about as well as ever.

1. I find it more difficult to make decisions than usual.

2. I have much greater difficulty in making decisions than I used to.

3. I have trouble making any decisions.

14. Worthlessness

0. I do not feel I am worthless.

1. I don't consider myself as worthwhile and useful as I used to.

2. I feel more worthless as compared to other people.

3. I feel utterly worthless.

15. Loss of Energy

0. I have as much energy as ever.

1. I have less energy than I used to have.

2. I don't have enough energy to do very much.

3. I don't have enough energy to do anything.

16. Changes in Sleeping Pattern

0. I have not experienced any change in my sleeping pattern.

la. I sleep somewhat more than usual.

lb. I sleep somewhat less than usual.

2a. I sleep a lot more than usual.

2b. I sleep a Iot less than usual.

3a. I sleep most of the day.

3b. I wake up 1-2 hours early and can't get back to sleep.

17. Irritability

0. I am no more irritable than usual.

1. I am more irritable than usual.

2. I am much more irritable than usual.

3. I am irritable all the time.

18. Changes in Appetite

0. I have not experienced any change in my appetite.

la. My appetite is somewhat less than usual.

lb. My appetite is somewhat greater than usual.

2a. My appetite is much less than before.

2b. My appetite is much greater than usual.

3a. I have no appetite

3b. I crave food all the time.

19. Concentration Difficulty

0. I can concentrate as well as ever.

1. I can't concentrate as well as usual.

2. It's hard to keep my mind on anything for very long.

3. I find I can't concentrate on anything.

20. Tiredness or Fatigue

0. I am no more tired or fatigued than usual.

1. I get more tired or fatigued more easily than usual.

2. I am too tired or fatigued to do a lot of the things I used to do.

3. I am too tired or fatigued to do most of the things I used to do.

21. Loss of Interest in Sex

0. I have not noticed any recent change in my interest in sex.

1. I am less interested in sex than I used to be.

2. I am much less interested in sex now.

3. I have lost interest in sex completely

at all.

2. Scoring instructions for BDI-II questionnaire

The depression level obtained from this questionnaire is regularly used to categorize users as: *minimal depression* (0-9), *mild depression* (10-18), *moderate depression* (19-29), and *severe depression* (30-63). A third method of evaluation will consist of assessing the systems in terms of how many users are correctly categorized (automatic questionnaire vs real questionnaire).

Appendix 2: Arabic Version of BDI questionnaire

يتكون هذا المقياس من 21 مجموعة من الأسئلة ، وكل مجموعة تصف أحد الأعراض السريرية للاكتئاب ، ويطلب من الشخص أن يقرأ كل عبارة من كل مجموعة ، وأن يقرر أي عبارة تنطبق عليه ، وتصف حالته ومشاعره ، بوضع دائرة حول رقم العبارة:

1. **الحزن**

0-لا أشعر بالحزن

1-أشعر بالحزن والكآبة.

2-الحزن والانقباض يسيطران علي طوال الوقت ، وأعجز عن الفكاك منهما.

3-أشعر بالحزن أو التعاسة لدرجة مؤلمة.

4-أشعر بالحزن والتعاسة لدرجة لا تحتمل.

2. **التشاؤم من المستقبل**

0-لا أشعر بالقلق أو التشاؤم من المستقبل.

1-أشعر بالتشاؤم من المستقبل.

2-لا يوجد ما أتطلع إليه في المستقبل.

3-لا أستطيع أبداً أن أتخلص من متاعبي.

4-أشعر باليأس من المستقبل ، وأن الأمور لن تتحسن.

3. **الإحساس بالفشل**:

0-لا أشعر بأني فاشل.

1-أشعر أن نصيبي من الفشل أكثر من العاديين.

2-أشعر أني لم أحقق شيئا له معنى أو أهمية.

3-عندما أنظر إلى حياتي في السابق أجدها مليئة بالفشل.

4-أشعر أني شخص فاشل تماماً (أباً أو زوجاً.)

4. **السخط و عدم الرضا**

0-لست ساخطاً .

1-أشعر بالملل أغلب الوقت.

2-لا أستمتع بالأشياء كما كنت من قبل.

3-لم أعد أجد شيئا يحقق لي المتعة (أو الرضا).

4-إنني غير راض وأشعر بالملل من أي شيء.

5. الإحساس بالندم أو الذنب

0-لا يصيبني إحساس خاص بالندم أو الذنب على شيء.

1-أشعر بأنني سيء أو تافه أغلب الوقت.

2-يصيبني إحساس شديد بالندم والذنب.

3-أشعر بأنني سيء وتافه أغلب الأوقات تقريباً.

4-أشعر بأنني سيء وتافه للغاية.

6. توقع العقاب

0-لا أشعر بأن هناك عقاباً يحل بي.

1-أشعر بأن شيئاً سيئاً سيحدث أو سيحل بي.

2-أشعر بأن عقاباً يقع علي بالفعل.

3-أستحق أن أعاقب.

4-أشعر برغبة في العقاب

7. كراهية النفس

0-لا أشعر بخيبة الأمل في نفسي.

1-يخيب أملي في نفسي.

2-لا أحب نفسي.

3-أشمئز من نفسي.

4-أكره نفسي.

8. إدانة الذات

0-لا أشعر بأني أسوأ من أي شخص آخر.

1-أنتقد نفسي بسبب نقاط ضعفي أو أخطائي.

2-ألوم نفسي لما أرتكب من أخطاء.

3-ألوم نفسي على كل ما يحدث.

9. وجود أفكار انتحارية

0-لا تنتابني أي أكار للتخلص من نفسي.

1-تراودني أفكار للتخلص من حياتي ولكن لا أنفذها.

2-أفضل لي أن أموت.

3-أفضل لعائلتي أن أموت.

4-لدي خطط أكيدة للانتحار.

5-سأقتل نفسي في أي فرصة متاحة.

10. البكاء

0-لا أبكي أكثر من المعتاد.

1-أبكي أكثر من المعتاد.

2-أبكي هذه الأيام طوال الوقت ولا أستطيع أن أتوقف عن ذلك.

3-كنت قادراً على البكاء ولكنني أعجز الآن عن البكاء حتى لو أردت ذلك.

11. الاستثارة وعدم الاستقرار النفسي

0-لست منزعجاً هذه الأيام عن أي وقت مضى.

1-أنزعج هذه الأيام بسهولة.

2-أشعر بالانزعاج والاستثارة دوماً

3-لا تثيرني ولا تغضبني الآن حتى الأشياء التي كانت تسبب ذلك سابقاً

12. الانسحاب الاجتماعي

0-لم أفقد اهتمامي بالناس.

1-أنا الآن أقل اهتماماً بالآخرين عن السابق.

2-فقدت معظم اهتمامي وإحساسي بوجود الآخرين.

3-فقدت تماما اهتمامي بالآخرين.

13. التردد وعدم الحسم

0-قدرتي على اتخاذ القرارات بنفس الكفاءة التي كانت عليها من قبل.

1-أؤجل اتخاذ القرارات أكثر من قبل.

2-أعاني من صعوبة واضحة في اتخاذ القرارات.

3-أعجز تماما عن اتخاذ أي قرار بالمرة.

14. تغير صورة الجسم والشكل

0-لا أشعر بأن شكلي أسوأ من قبل.

1-أشعر بالقلق من أني أبدو أكبر سناً وأقل جاذبية.

2-أشعر بوجود تغيرات دائمة في شكلي تجعلني أبدو منفراً (منفرة) وأقل جاذبية.

3-أشعر بأن شكلي قبيح (قبيحة) ومنفر (منفرة.(

15. هبوط مستوى الكفاءة والعمل

0-أعمل بنفس الكفاءة كما كنت من قبل.

1-أحتاج إلى مجهود خاص لكي أبدأ شيئاً

2-لا أعمل بنفس الكفاءة التي كنت أعمل بها من قبل.

3-أدفع نفسي بمشقة لكي أعمل أي شيء.

4-أعجز عن أداء أي عمل على الإطلاق

16. اضطرابات النوم

0-أنام جيداً كما تعودت.

1-أستيقظ مرهقاً في الصباح أكثر من قبل.

2-أستيقظ من 2-3 ساعات أبكر من ذي قبل ، وأعجز عن استئناف نومي.

3-أستيقظ مبكراً جداً ولا أنام بعدها حتى إن أردت.

17. التعب والقابلية للإرهاق

0-لا أتعب بسرعة أكثر من المعتاد.

1-أشعر بالتعب والإرهاق أسرع من ذي قبل.

2-أشعر بالتعب حتى لو لم أعمل شيئا.

3-أشعر بالتعب الشديد لدرجة العجز عن عمل أي شيء.

18. فقدان الشهية

0-شهيتي للطعام ليست أسوأ من قبل.

1-شهيتي ليست جيدة كالسابق.

2-شهيتي أسوأ بكثير من السابق.

3-لا أشعر برغبة في الأكل بالمرة.

19. تناقص الوزن

0-وزني تقريباً ثابت.

1-فقدت أكثر من 3 كغ من وزني.

2-فقدت أكثر من 6 كغ من وزني.

3-فقدت أكثر من 10 كغ من وزني.

20. تأثر الطاقة الجنسية

0-لم ألاحظ أي تغيرات حديثة في رغبتي الجنسية.

1-أصبحت أقل اهتماماً بالجنس من قبل.

2-قلت رغبتي الجنسية بشكل ملحوظ.

3-فقدت تماما رغبتي الجنسية.

21. الانشغال على الصحة

0-لست مشغولاً على صحتي أكثر من السابق.

1-أصبحت مشغولاً على صحتي بسبب الأوجاع والأمراض ، أو اضطرابات المعدة والإمساك.

2-أنشغل بالتغيرات الصحية التي تحدث لي لدرجة أني لا أستطيع أن أفكر في أي سيء آخر.

3-أصبحت مشغولاً تماما بأموري الصحية

مفتاح التصحيح

0-9	لا يوجد اكتئاب
10-15	اكتئاب بسيط
16-23	اكتئاب متوسط
24-36	اكتئاب شديد
37 فأكثر	اكتئاب شديد جدًا

Appendix 3: DSM-5

DSM-5 Diagnostic Criteria for Major Depressive Disorder
A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure. **Note**: Do not include symptoms that are clearly attributable to another medical condition.
1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful). (**Note**: In children and adolescents, can be irritable mood.)
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).
3. Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month) or decrease or increase in appetite nearly every day. (**Note**: In children, consider failure to make expected weight gain.)
4. Insomnia or hypersomnia nearly every day.
5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).
6. Fatigue or loss of energy nearly every day.
7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).
9. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.
B. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
C. The episode is not attributable to the physiological effects of a substance or to another medical condition.
Note: Criteria A–C represent a major depressive episode.

Note: Responses to a significant loss (e.g., bereavement, financial ruin, losses from a natural disaster, a serious medical illness or disability) may include the feelings of intense sadness, rumination about the loss, insomnia, poor appetite, and weight loss noted in Criterion A, which may resemble a depressive episode. Although such symptoms may be understandable or considered appropriate to the loss, the presence of a major depressive episode in addition to the normal response to a significant loss should also be carefully considered. This decision inevitably requires the exercise of clinical judgment based on the individual's history and the cultural norms for the expression of distress in the context of loss.
D. The occurrence of the major depressive episode is not better explained by schizo-affective disorder, schizophrenia, schizophreniform disorder, delusional disorder, or other specified and unspecified schizophrenia spectrum and other psychotic disorders.
E. There has never been a manic episode or a hypomanic episode. Note: This exclusion does not apply if all of the manic-like or hypomanic-like episodes are substance-induced or are attributable to the physiological effects of another medical condition.
DSM = *Diagnostic and Statistical Manual of Mental Disorders*. *Reprinted with permission from the American Psychiatric Association.* Diagnostic and Statistical Manual of Mental Disorders, *5th ed. Washington, DC: American Psychiatric Association; 2013:160-161.*

الملّخص

يُعدّ مسح الاكتئاب السريري من أكبر التحديات الّتي تواجه الأطباء النفسيين، وذلك لارتفاع التكاليف المترتبة عليه إضافةً إلى الوصمة المجتمعية المرتبطة بالأمراض النفسية والّتي تحول دون تلقي الرعاية النفسية اللازمة. تشير منظمة الصحة العالمية إلى أنّ الاكتئاب السريري مرض شائع يصيب أكثر من 264 مليون شخص حول العالم ويقدّر تأثيره على الاقتصاد العالمي السنوي بنحو 1 تريليون دولار من الانتاجية المفقودة. لذلك، فإنّ استخدام مقاييس تشخيصية ذات دقة في عملية المسح للاكتئاب السريري يعدّ جزءًا أساسيًا من عملية توفير الرعاية النفسية الملائمة. ومن بين أدوات القياس ذاتية الاستخدام المتوفرة مقياس بيك الثاني للاكتئاب. ويعدّ مقياس بيك الثاني للاكتئاب -والذي يتكوّن من 21 فئة- من أكثر المقاييس شيوعًا نظرًا لدقته في تمييز الأشخاص المكتئبين عن غير المكتئبين وتحديد شدّة الاكتئاب لديهم. وقد وضعت هذه الفئات في العام 1996 وتمت مراجعتها وفقًا لمعايير التشخيص التابعة للدليل التشخيصي والإحصائي الخامس للاضطرابات العقلية. وتعتمد فئات مقياس بيك الثاني للاكتئاب والتي لم يتم التعديل عليها منذ ذلك الحين على مقاييس ذاتية ومترابطة لغويًا بدلًا من استخدام فئات مستقلة رياضيًا، ولذلك فإنّ دقة هذه الفئات غير المنظمة محدودة للغاية. وللمساعدة في تقليل الفجوة العلاجية لمرضى الاكتئاب السريري، أصبح من المهم استخدام الأساليب التكنولوجية لتطوير أدوات مسح جديدة مبنية على الأسئلة الّتي وضعها الأطباء النفسيين سابقًا. توفّر تقنيات معالجة اللغة الطبيعية أداة فحص للتنبؤ بأعراض الاكتئاب السريري وشدّته. وفي هذا المشروع البحثي نهدف إلى استخدام معالجة اللغة الطبيعية لبناء شبكة دلالية معجمية تُثري الأسئلة المستخدمة في مقياس بيك الثاني للاكتئاب باستخدام العلاقات المستنبطة من *Wordnet* وشبكات التواصل الاجتماعي ونظام اللغة الطبية الموحد *UMLS*. ويتمثل هذا الإثراء من خلال إعادة صياغة الأسئلة الأصلية وإعادة ترتيبها باستخدام العلاقات المشتقة لغويًا من مصادر خارجية. ومن ثمّ، تطوير أداة فحص ثنائية اللغة باللغتين العربية والإنجليزية لتوفير مسح أكثر دقة لمرض الاكتئاب السريري والمساعدة في توفير العلاج والرعاية النفسية والبدنية المطلوبين. ويراعي تصميم الأداة المنتجة إمكانية استخدامها من قبل المرضى بشكل ذاتي دون الحاجة لرؤية الطبيب؛ بينما تقدّم الشبكة المعجمية المُخرَجة أداة فورية الاستخدام من قبل الباحثين الآخرين الذين يدرسون أعراض الاكتئاب السريري وتأثيره على المريض. وبمقارنة الأداة المطورة في هذا البحث مع الفئات المستخدمة حاليًا، فإنّ هذه الأداة

تهدف إلى تسريع عملية المسح وجعلها أكثر مصداقية من خلال استخدام الفئات الدلالية المناسبة للاكتئاب السريري. كما وتهدف الأداة إلى زيادة دقة عملية المسح من خلال إعادة تجميع الأسئلة في فئات تعبّر عن نفس الدلالات الطبية بناءً على نماذج المجموعات المتعددة.