الجمـهورية الجزائرية الديمقراطية الشعبيـة
وزارة التعليم العالي والبحث العلمي
جـامعـة سعـيـدة د. مـولاي الطـاهـر
كلية الرياضيات و الإعلام الآلي و الاتصالات السلكية و اللاسلكية
قسم: الإعلام الآلي

Mémoire de Master en informatique

Spécialité : Intelligence Artificielle Principe et application

T h è m e

Mental health diagnosis through text analysis

▪ **Présenté par :**
Ghouti Farah Radhia

▪ **Dirigé par :**
Mme Kadari Rekia

Année universitaire 2024-2025

Acknowledgements

First and foremost, I would like to thank Allah for the strength, knowledge, and opportunity he has provided me to accomplish this thesis.

I would also like to express my appreciation and gratitude to my supervisor Dr.KADARI Rekia, for her availability, patience, and valuable advice.
I also want to thank all my professors in the Computer Science Department who, directly or indirectly, contributed to the completion of this work.
A big thank you to my family, who have supported and encouraged me throughout all these years, as well as to my friends for their support and assistance.

Contents

List of Figures

List of Tables

List of Tables

Introduction

Our daily well-being is influenced by our thoughts, feelings, and actions. No one is ever truly content, thinks positively, and acts rationally all the time. Nonetheless, it is usually lot simpler to appreciate life and deal with its difficulties when we are in a generally positive frame of mind. Although taking care of our mental health is equally as important as taking care of our physical health, most of us are much better at managing our physical health than our mental health. We usually strive to take action as soon as we experience bodily discomfort or ache, but when we are feeling depressed or anxious, we often assume that this is a normal part of life and do nothing to make things better [2].

Mental disorders become a pressing challenge in today's society. The World Health Organization (WHO) highlights that mental health disorders can significantly impact all aspects of life, including personal relationships, daily functioning and even the academic performance .WHO also reported that individuals with severe mental health disorders die 10 to 20 years earlier than the general population. Furthermore, mental disorders considerably increase the risk of suicide [3]. However, significant gaps persist in mental health care. Traditional therapeutic services are often costly and inaccessible for many, which, combined with societal stigma, discourages individuals from seeking help. Additionally, a large portion of the population lacks awareness or understanding of their own mental health conditions, leading to delayed diagnosis and treatment. This underscores the urgent need for affordable, stigma-free, and easy-to-use tools that can assist individuals in recognizing early signs of mental health issues and seeking timely support.

As the prevalence of mental health problems increases, there is an increasing need for innovative solutions to support both diagnosis and treatment using artificial intelligence. Sequence labeling corresponds to several tasks of natural language processing (NLP) that aim at assigning a sequence of symbols (i.e. words) with labels. Many of the early pioneers of CL researchers were interested in the area of sequence labeling since it seems so useful for many tasks. In the last few decades, mental health disorders detection attracts the attention of several researchers and become more and more important in the field of NLP.

Mental health disorders detection can be classified as a sequence labelling task where each sentence (or paragraph or speech segment) is assigned a corresponding label. In literature, dominant approaches based on machine learning methods have been proposed for the mental health disorders detection task such as Random Forest and Linear Regression. However, the use of machine learning in NLP has been mostly limited to numerical optimization of weights for humanly designed representations and features from the text data.

In this thesis, we focus on the automated mental health disorders detection in order to propose and develop some techniques, which allow an early detection of mental disorders such as anxiety and depression. Our goal is the development of simple and accurate models that can solve the

challenging problem of multi classification and learn the necessary intermediate representations of input entries without the need for extensive features engineering based on deep learning models. This thesis is organized in four chapters. The first chapter introduces a brief overview of Mental health disorders. Next, chapter 2 provides a brief review of the previously machine learning proposed models for mental health disorders. In addition, Chapter 3 introduces the proposed deep learning models architectures. Finally, the results of all the achieved results are discussed in chapter 4.

Chapter 1

Mental health disorder

Mental health disorders have become a major global concern, according to the World Health Organization (WHO). In 2019, approximately one in eight people around 970 million individuals worldwide lived with a mental disorder [4]. The number has increased significantly due to the COVID-19 pandemic, poor lifestyle choices, sleep deprivation, social isolation, and sedentary behavior. Most people with mental health disorders lack access to quality care, despite the existence of effective prevention and treatment options. Additionally, many people encounter stigma, discrimination, and human rights violations [4].

1.1 Mental health disorder

"Mental disorder is characterized by a clinically significant disturbance in an individual's cognition, emotional regulation, or behavior. It is usually associated with distress or impairment in important areas of functioning. There are many different types of mental disorders. Mental disorders may also be referred to as mental health conditions. The latter is a broader term that covers mental disorders, psychosocial disabilities, and (other) mental states associated with significant distress, impairment of functioning, or risk of self-harm." [4] Mental health disorder, also known as mental illness, is a form of emotional, cognitive, or behavioral dysfunction that affects the brain and alters its functions.It is not a short-term feeling but a persistent condition over time. Mental disorders can originate from biological, psychological, or environmental factors and often reduce an individual's quality of life.

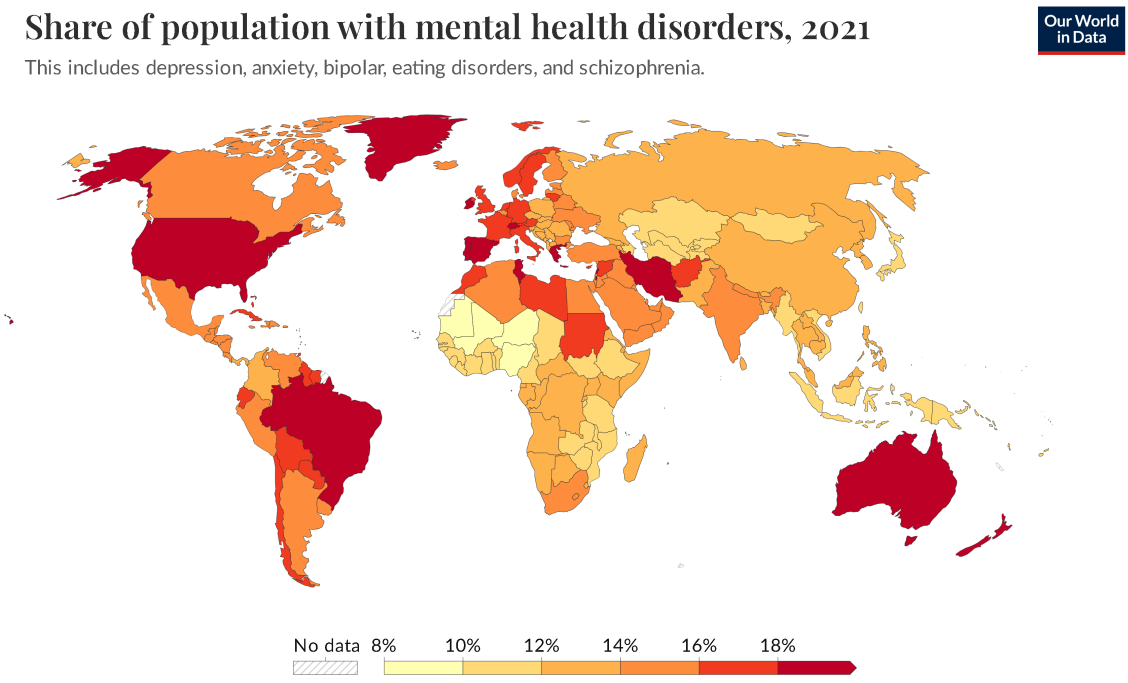


Figure 1.1: Global share of people with mental and substance use disorders (2021). Source: Our World in Data.

1.2 Mental health disorders classes

1.2.1 Depression

"Depression is a mood disorder that causes a persistent feeling of sadness and loss of interest. Also called major depressive disorder or clinical depression, it affects how you feel, think, and behave and can lead to a variety of emotional and physical problems. You may have trouble doing normal day-to-day activities, and sometimes you may feel as if life is not worth living." [5].

1.2.2 Depression symptoms

- Feelings of sadness, tearfulness, emptiness or hopelessness
- Angry outbursts, irritability or frustration, even over small matters
- Sleep disturbances, including insomnia or sleeping too much
- Tiredness and lack of energy, so even small tasks take extra effort
- Reduced appetite and weight loss or increased cravings for food and weight gain
- Anxiety, agitation or restlessness
- Unexplained physical problems, such as back pain or headaches
- Trouble thinking, concentrating, making decisions ,and remembering things

1.2.3 Anxiety

"Anxiety is usually a natural response to pressure, feeling afraid or threatened, which can show up in how we feel physically, mentally, and in how we behave. It is common to describe anxiety as a feeling of dread, fear, or unease, which can range from mild to severe. " [6].

1.2.4 Anxiety symptoms

- Persistent worrying or anxiety about a number of areas that are out of proportion to the impact of the events
- Overthinking plans and solutions to all possible worst-case outcomes
- Fatigue

- Trouble sleeping
- Muscle tension or muscle aches
- Nervousness or being easily startled
- Nausea, diarrhea or irritable bowel syndrome
- Difficulty concentrating, or the feeling that your mind "goes blank"

1.2.5 ADHD

"Attention-deficit/hyperactivity disorder is one of the most common mental disorders affecting children. Symptoms of ADHD include inattention (not being able to keep focus), hyperactivity (excess movement that is not fitting to the setting) and impulsivity (hasty acts that occur in the moment without thought)" [7].

1.2.6 ADHD symptoms

- Disorganization and problems prioritizing
- Problems focusing on a task
- Trouble multitasking
- Excessive activity or restlessness
- Frequent mood swings
- Problems following through and completing tasks
- Trouble coping with stress
- Poor time management skills

1.2.7 PTSD

"Post-traumatic stress disorder is a psychiatric disorder that may occur in people who have experienced or witnessed a traumatic event, series of events or set of circumstances. An individual may experience this as emotionally or physically harmful or life-threatening and may affect mental, physical, social, and / or spiritual well-being." [8].

1.2.8 PTSD symptoms

PTSD symptoms are grouped into four types:

Intrusive memories

- Unwanted, distressing memories of a traumatic event that come back over and over again.
- Upsetting dreams or nightmares about a traumatic event.
- Severe emotional distress or physical reactions to something that reminds you of a traumatic event.

Avoidance

- Trying not to think or talk about a traumatic event.
- Staying away from places, activities, or people that remind you of a traumatic event.

Negative changes in thinking and mood

- Negative thoughts about yourself, other people, or the world.
- Feeling detached from family and friends.

Changes in physical and emotional reactions

- Always being on guard for danger.
- Physical reactions, such as sweating, rapid breathing, heartbeat, or shaking.

1.2.9 Bipolar

"Bipolar disorder (formerly called manic-depressive illness or manic depression) is a mental illness that causes clear shifts in a person's mood, energy, activity levels, and concentration. People with bipolar disorder often experience periods of extremely 'up', irritable, or energized behavior (known as manic episodes) and very 'down', sad, indifferent, or hopeless periods (known as depressive episodes)." [9].

1.2.10 Bipolar symptoms

Bipolar disorder includes two episodes, each associated with different symptoms.

Mania and hypomania

- Being much more active, energetic or agitated than usual
- Feeling a distorted sense of well-being or too self-confident
- Needing much less sleep than usual
- Being unusually talkative and talking fast

Major depressive episode

- Feeling restless or acting slower than usual
- Being very tired or losing energy
- Having a hard time thinking or concentrating, or not being able to make decisions
- Thinking about, planning or attempting suicide

All symptoms listed in this section were taken from Mayo Clinic [10–14].

1.3 Causes of mental health disorders

Many genetic and environmental factors are believed to be responsible for mental illnesses :

1.3.1 Inherited traits

People with blood relatives who have mental health illnesses are more likely to suffer from them. Certain genes may increase the risk of developing a mental disorder and life situations may trigger it [15].

1.3.2 Environmental exposures before birth

Mental illness can sometimes be connected to prenatal exposure to environmental stressors,drugs,toxins, inflammatory conditions and alcohol [15].

1.3.3 Brain chemistry

A naturally occurring brain chemical substances, called neurotransmitters , transmit signals between various regions of the body and brain. When the neural networks involving these chemicals are disrupted , the functioning of nerve receptors and neural systems is altered, leading to depression and other emotional disorders [15].

1.4 Diagnosis

In order to establish a diagnosis and check for any associated complications, a physical exam,laboratory tests, and a psychological evaluation may be performed:

1.4.1 Physical exam

A Physical assessment is conducted by a health care provider in order to rule out any physical issues that may be causing symptoms such as cardiovascular disease, asthma, or diabetes [16].

1.4.2 Lab tests

Laboratory examinations are performed, including, for instance a thyroid function test, an alcohol test or drug substances use . psychological [16].

1.4.3 A psychological evaluation

Feeling, behavior patterns, and thoughts will be discussed with a mental health specialist.They may ask patients to fill out a questionnaire to assist in answering questions [16].

1.5 Treatments

The nature and the severity of mental illness, along with what is most effective in each case , determine the appropriate treatment. A mix of treatments is often the most effective.

A primary care physician may be able to manage a moderate mental illness if the symptoms are under control. However, to ensure that all social, medical, and mental requirements are satisfied, a team approach is frequently suitable. For serious mental diseases like schizophrenia, such coordination is particularly crucial [17].

1.5.1 Types of Treatment

Medications

The following are a few of the most widely prescribed classes of mental health medications.

- Antidepressants
- Anti-anxiety medications
- Mood-stabilizing medications
- Antipsychotic medications

Psychotherapy

Talking with a mental health expert about your ailment and associated problems is known as psychotherapy, or talk therapy. You gain knowledge of your condition as well as your emotions, thoughts, behavior, and moods during psychotherapy. You can develop coping and stress management techniques using the information and understanding you acquire [17].

1.5.2 Brain-stimulation treatments

Treatments for depression and other mental health conditions occasionally involve brain stimulation. They are often saved for cases where psychotherapy and medication have failed. These consist of vagus nerve stimulation, deep brain stimulation, repetitive transcranial magnetic stimulation, and electroconvulsive therapy [17].

1.5.3 Hospital and residential treatment programs

Mental disease might occasionally worsen to the point that you require treatment in a mental health facility. When you are in imminent danger of hurting yourself or someone else, or when you are unable to properly take care of yourself, this is usually recommended [17].

1.6 Prevention strategies

Mental illness cannot be completely avoided. Nonetheless, if you suffer from a mental illness, managing stress, building resilience, and improving poor self-esteem can help manage your symptoms. Take these actions:

Be mindful of warning indicators. Collaborate with your physician or therapist to identify potential triggers for your symptoms. Plan beforehand so you know what to do in the event that symptoms reappear. If your symptoms or feelings change, get in touch with your physician or therapist. Think about asking friends or relatives to keep an eye out for any warning indications.
Seek regular medical attention. Avoid skipping appointments or checkups with your primary care physician, particularly if you are feeling under the weather. You can be dealing with a new health issue that requires treatment, or you might be suffering adverse drug reactions.
When you need assistance, get it. Waiting until symptoms worsen can make treating mental health disorders more difficult. A symptom recurrence may also be avoided with long-term maintenance treatment.
Finally, Look after yourself. It is crucial to get enough sleep, eat well, and exercise frequently. Make an effort to keep a consistent routine. If you have concerns about your diet or level of physical activity, or if you have problems falling asleep, speak with your health care physician [18].

1.7 Conclusion

Mental health disorders are complex, multi-factorial conditions that affect emotional, behavioral, and cognitive well-being. Despite the availability of effective treatments, lack of awareness and limited health care access remain major obstacles. Understanding types, causes and symptoms is essential for accurate diagnosis and early prevention. Addressing the impact of these disorders is crucial in reducing their burden on society, individuals, and overall quality of life.

Chapter 2

Machine Learning for Mental Health Disorders

> **"Artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment."**
> *—Nils J. Nilsson, AI Definition*

Artificial intelligence (AI) lacks a single widely agreed on definition of the AI concept. This is due to its complex, wide, and interdisciplinary nature, including, among other domains, computer science, mathematics, psychology, linguistics, and philosophy.Some define AI in the broadest sense, equating it to algorithms. While others define it in a more specific definition as the ability of computers to mimic human intelligence [19].For instance, John McCarthy, known as the father of AI, defines it as

> **"Artificial intelligence is the science and engineering of making intelligent machines. "**

2.1 Artificial Intelligence

Although various definitions exist, AI can also be defined as a system that enables machines to reason and adapt based on data, environment, and experiences and provides solutions by mimicking human cognitive processes without human involvement.

Recently, the application of AI has expanded across nearly all areas. Among these, the mental health care domain, as AI is changing the way mental health conditions are identified, treated, and managed. It allows for personalized treatment and better interaction, in addition to improving

early detection by analyzing user data by applying natural language processing (NLP) techniques and machine learning algorithms [20].

2.2 Machine Learning

A wide variety of algorithms that make intelligent predictions using a collection of data are together referred to as machine learning (ML). These data sets are frequently sizable; they may include millions of distinct data points. Recent advances in machine learning have achieved what seems to be a human level of information extraction and semantic understanding, and occasionally the capacity to identify abstract patterns more accurately than human specialists [21].

Modern machine learning, which builds on traditional statistical modeling techniques, has become a powerful tool as a result of exponentially growing processing power, significantly expanded data volumes,and improvements in algorithm design.

Today, there are many different types of machine learning algorithms. The type of desired result and the data's features indicate which model is best suited for a given challenge. The quantity of distinct data points is one of the main factors. More complex deep learning methods might be appropriate for large data sets. A smaller number of data points suggests that reliable traditional methods such as linear regression, or decision-tree techniques, which divide data sets into areas based on predetermined guidelines, are probably going to work better. Whether the data is a time-series signal, a collection of photos, or generic descriptive data, care must be taken to customize the technique to the specifics of the data [21].

2.3 Deep Learning

One type of machine learning that works significantly better with unstructured data is called deep learning. Present-day machine learning methods are being surpassed by deep learning methods. It makes it possible for computer models to gradually extract features from data at various levels. With the growth of data and the development of hardware that produced powerful computers, deep learning became more and more popular [22]. Most machine learning based models had exploited shallow structured architectures. These architectures typically contain at most one or two layers. Gaussian Mixture Models (GMMs), linear or nonlinear dynamical systems, CRFs, ME models, Support Vector Machines (SVMs), and Multi-Layer Perceptron (MLP) are some examples of the shallow architectures. Despite the effectiveness of shallow architectures to solve many simple or strained problems; their main disadvantages are their limited modeling and representational power which can cause difficulties when dealing with more complicated real-world applications. Moreover, those methods require the design and selection of an appropriate feature space to be developed by experts and it is costly, and difficult in terms of computational time or expert knowledge. As an

alternative, automatically learning the features can be considered as a relevant choice. Artificial Neural Networks (ANNs) models have been introduced over decades if not centuries. Earlier studies with ANNs were started in the late 1950s with the introduction of the perceptron, a two-layer network used for simple operations, and growth in the late 1960s with the development of an efficient gradient descent method called the back-propagation algorithm [23]. applied to NN for efficient multilayer networks training. ANNs represent a class of machine learning models. In ANNs, the artificial neuron forms the computational unit of the model and the network describes how these units are connected to one another. The simplest version of ANNs is feed-forward Neural Network(FNN). Basically, a FNN receives and maps a set of inputs to outputs. Each NN is constructed by several interconnected neurons, organized in layers with associated weights. An example of a FNN is shown in Figure 2.1 which consists of three layers: the input layer which read inputs and transfer them to the hidden layer performing computations to be transferred to the output layer.

FNNs have been highly promising in the field of mental health prediction, particularly in the detection of depression and other psychiatric conditions, without the need for manual feature extraction [24]. The significant disadvantage in such networks is the large number of free parameters (weights) to be optimized. In addition, FNNs are constrained by their lack of ability to model temporal dependencies because inputs are processed in one direction. Thus, the output at every time step relies solely on the current input, without any information about the neighboring inputs.

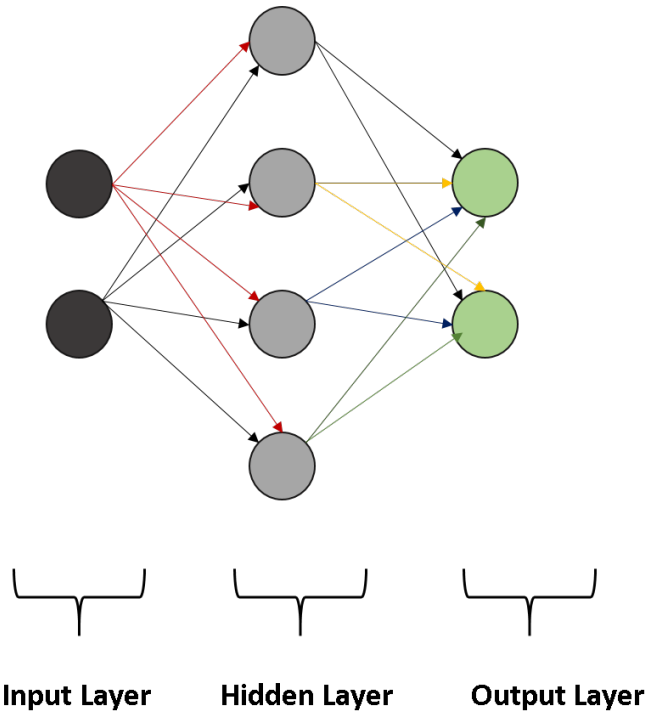


Figure 2.1: An example of an Artificial Neural Network.

Since the mid-2000s, there has been renewed interest in neural networks, marked by the invention of a fast-learning algorithm by G. Hinton [25], and the widespread use of GPUs for large-scale numerical calculations around 2011. These advances enabled the advent of modern deep learning, which is characterized by deeper and better-performing neural network architectures, as demonstrated in Figure 2.2.

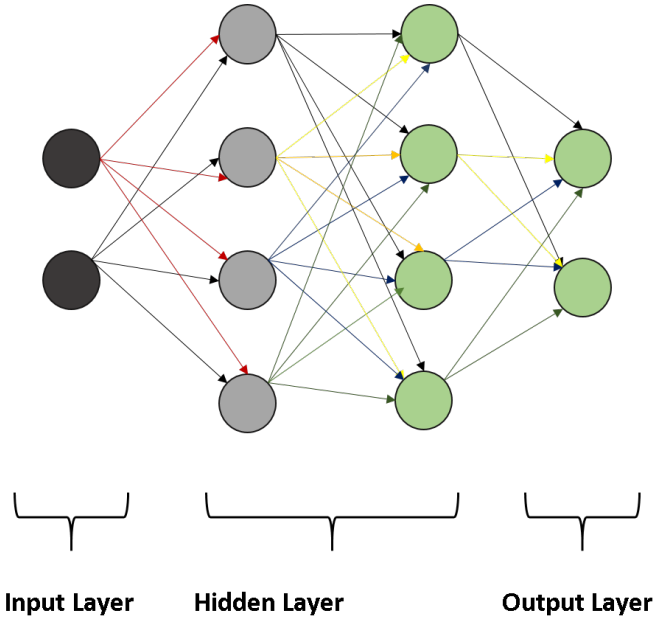


Figure 2.2: An example of a Deep Neural Network.

FNN or Multi-Layer Perceptrons MLPs with many hidden layers, often

referred to deep neural networks (DNNs), are examples of the models with a deep architecture with the presence of more than two layers, an input layer, one or more so-called ”hidden” layers, and an output layer. A few years ago, researchers called the deep learning networks as ”deep” with 3-5 layers, and now it has gone up to 100-200 layers. Deep Learning has appeared as the new area of machine learning research [25] [26], with the objective of moving machine learning towards its original goals. Modern deep learning networks have been applied with success for many complete problems. By adding more levels (layers), researchers reported positive experimental results for several tasks [27] [28] [29] [30]. Since the mid of the 2000s to nowadays, the techniques developed from deep learning research have already been impacting a wide range of information processing work within the traditional and the new, key aspects of machine learning and AI [26] [31] [32] [33] [34].

2.4 Natural Language Processing

”According to IBM Natural language processing (NLP) is a subfield of computer science and artificial intelligence (AI) that uses machine learning to enable computers to understand and communicate with human language.NLP enables computers and digital devices to recognize, understand and generate text and speech by combining

computational linguistics, the rule-based modeling of human language together with statistical modeling, machine learning and deep learning".

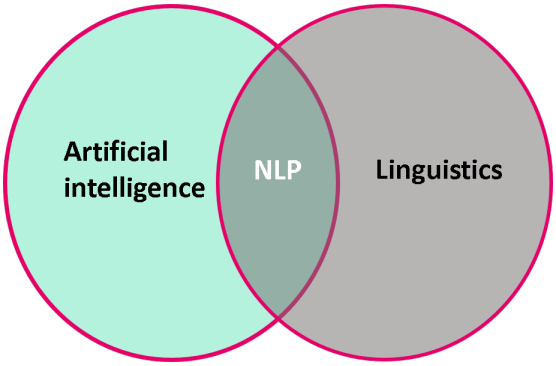


Figure 2.3: Venn diagram showing relationship between NLP and AI

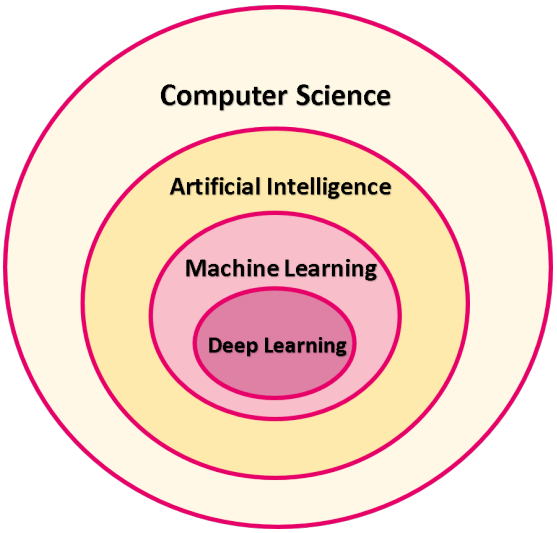


Figure 2.4: Venn diagram showing relationship between CS, AI, ML, and DL

2.5 NLP subdomains

2.5.1 Part-Of-Speech-Tagging

A portion of speech The practice of giving each word in a document a part-of-speech is known as tagging. A sequence of (tokenized) words and a tagset

$$(x_1, x_2, \ldots, x_n)$$

is the input, and a sequence is the output.

$$(y_1, y_2, \ldots, y_n)$$

of tags, with each output

$$(y_i)$$

precisely matching one input

$$(x_i)$$

as illustrated in Figure 2.5 [35].

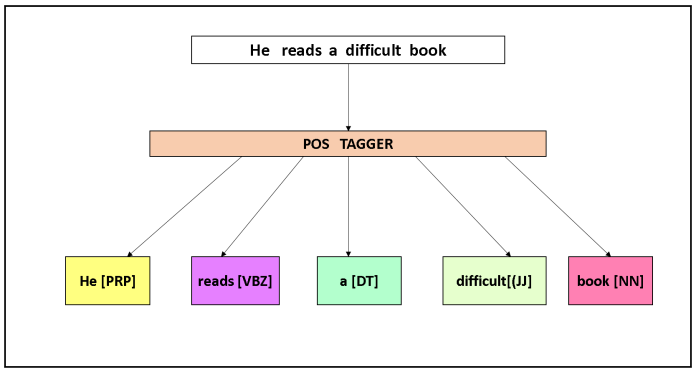


Figure 2.5: An illustration of the POS tagging process.

2.5.2 Information Extraction

Information Extraction (IE) The unstructured information contained in texts is transformed into structured data by the information extraction (IE) process, which is used, for instance, to fill a relational database and facilitate additional processing [35].

2.5.3 Information Retrieval

"Information Retrieval (IR) Information retrieval or IR is the name of the field encompassing the retrieval of all information retrieval IR manner of media based on user information needs. The resulting IR system is often called a search engine.The IR task we consider is called ad hoc retrieval, in which a user poses a query to a retrieval system, which then returns an ordered set of documents from document some collection. A document refers to whatever unit of text the system indexes and retrieves (web pages, scientific papers, news articles, or even shorter passages like collection paragraphs). A collection refers to a set of documents being used to satisfy user term requests. A term refers to a word in a collection, but it may also include phrases. query Finally, a query represents a user's information need expressed as a set of terms" [35].

2.5.4 Named Entity Recognition

In general, everything that has a proper name—such as a person, place, or organization—is referred to as a named entity (NE). Named entity recognition (NER) is the process of identifying textual segments that make up proper names and labeling the entity with the named entity recognition type. PER (person), LOC (location), ORG (organization), and GPE (geo-political entity) are the four most often used entity tags. Dates, timings, and other temporal expressions, as well as

numerical expressions like prices, are examples of items that are frequently included under the term "named entity" even though they are not entities in and of themselves [35].

2.5.5 Emotion Detection

Textual Emotion Detection (TED) is an important application of Natural Language Processing (NLP) facilitating automated analysis and detection of the primary emotion expressed in text. While a more general NLP application of sentiment analysis usually focuses on the coarse-grained classification of text into positive, neutral, or negative polarity, fine-grained analysis using TED generally uses scales of emotional labels, such as the six basic emotions introduced by Ekman : sadness, anger, surprise, fear, happiness, and disgust [36].

Emotion detection may improve a range of language processing tasks. Emotion recognition may be used by dialogue systems, such tutoring systems, to determine a student's moods. By automatically recognizing emotions (such anger, discontent, and trust) in reviews or consumer interactions, businesses may be able to pinpoint specific problem areas or ones that are succeeding.In medical NLP tasks, such as diagnosing depression or suicidal intent, emotion can be important. Understanding how various socioeconomic groupings were perceived by society at various points in time may be aided by identifying the feelings that characters in novels exhibit [35].

2.5.6 Machine Translation

Machine Translation (MT) is the process of translating text from one language to another using computers [35].As shown in Figure 2.6. The goal of machine translation in NLP is to provide translations that accurately capture the original content's meaning while also preserving the grammatical structure.

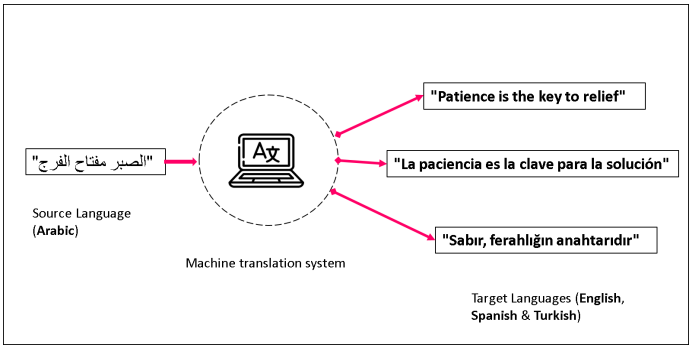


Figure 2.6: Illustration of machine translation: sentence-level translation from English to Spanish, French, and Turkish.

2.5.7 Question Answering System

Question Answering (Q&A) systems aim to satisfy human information needs by providing precise answers to natural language questions.

Answering questions is strongly related to the function of search engines because a lot of information is available in text format (on the internet or in other data like our emails or books). In fact, as massive language models trained to answer questions are incorporated into contemporary search engines, the line between the two is becoming increasingly hazy. A helpful subset of information demands factual questions—questions of fact or logic that can be addressed with straightforward facts presented in brief or medium-length texts—are frequently the subject of question-answering systems [35].

2.5.8 Text Classification

Classifying a set of N documents is known as the Text Classification (TC) process. To begin, we construct a classifier T. Next, we have a collection of D text documents, and each text document is assigned a class or label by an expert. Second, we must use a matching set of documents in D as input to train a classifier for each class or label. The trained classifier C must now be used to categorize N documents. Every document in N will be given a predetermined class or label by C, As presented in Figure 2.7. Text classification is a thorough procedure that involves a number of additional procedures, including as data pretreatment, transformation, and dimensionality reduction, in addition to model training [37].

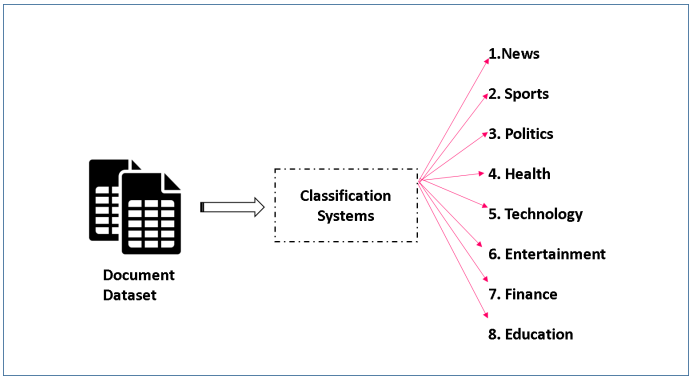


Figure 2.7: Text classification in NLP

2.5.9 Sentiment Analysis

The technique of thoroughly examining data available on the Internet in order to recognize and classify the views expressed d in a passage of text is known as sentiment analysis, or SA. This procedure is to evaluate the author's attitude toward a specific subject, film, product, etc. The outcome can be neutral, bad, or favorable. As illustrated in Figure 2.8. These studies demonstrated various SA strategies for analyzing and extracting feelings related to the polarity of positive, negative, or neutral on the chosen themes. Social media platforms SA can provide valuable statistics and information. SA is significant in a variety of business, political, and intellectual spheres [38].

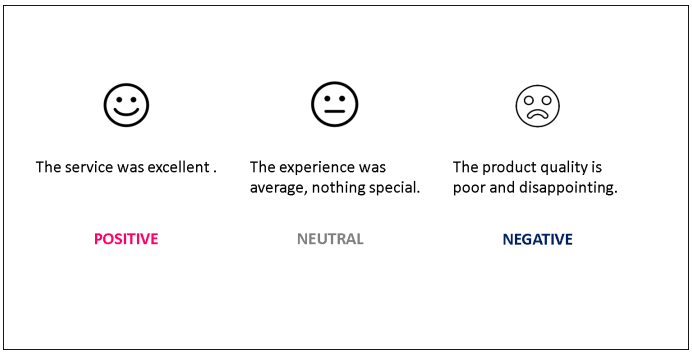


Figure 2.8: Sentiment Analysis in NLP

2.5.10 Text Summarization

In line with IBM For better information extraction, text summarizing reduces one or more texts into concise summaries.

One natural language processing (NLP) technique that distills information from one or more input text documents into an original output text is automatic text summarizing, also known as document summarization. There is disagreement about how much of the input text is included in the output; some definitions indicate 10%, while others state 50%. In order to read texts and produce their summaries, text summarization algorithms frequently make use of deep learning architectures, particularly transformers [35].

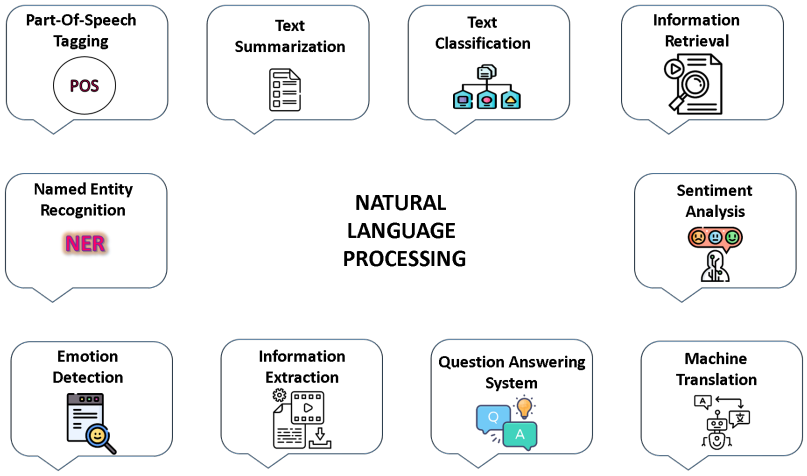


Figure 2.9: Some Subdomains of Natural Language Processing

2.6 Machine Learning proposed methods for mental health disorder diagnosis

The area of mental health analysis using NLP has been enriched over the last few decades by the contribution from several researchers. Many new techniques have been introduced to improve the efficiency of mental health detection.

More recently, several models have been used for the task of mental health analysis.Several advanced machine learning algorithms have been developed that acquire more robust information from textual data. In general, all machine learning models rely on hand-crafted features to provide good results, recent research has increasingly focused on deep learning models to address the limitations of manual feature extraction.Finally, combinations of several machine learning and deep learning models have been used in the current research direction.

In this section ,we will discuss some of prior studies in the field of mental health analysis using machine learning and deep learning approaches.

2.6.1 Detecting Mental Health Disorders Based on Social Media Data

Ankit Murarka,et al [39]. conducted a study focused on classifying mental illness using social media posts. They collected dataset of **17159** entries (posts+titles) through Reddit API, targeting only five specific mental health disorders.They used LSTM as a baseline model, alongside BERT and RoBERTa. The best F1-score was achieved with RoBERTa , compared to LSTM and BERT. The authors highlighted the key limitation of the study is its reliance on Reddit content, which may not be applicable to other platforms .Additionally the the task addressed only five disorders. The researchers aim to create a multi-label dataset and extend the classification to include more mental disorders.

2.6.2 Deep and Transfer Learning for Mental Disorders Detection in Social Media Posts

Building on the previous work, Muhammad Arif, et al, developed a mental health classification models using the same dataset as the first article [40].they applied several machine learning algorithms including Random Forest (RF),linear support vector machine (SVM), Multinomial Naive Bayes (NB), and Logistic Regression (LR), Among these, Logistic Regression achieved the highest F1-score using word n-grams features.
Alongside they also explored deep learning models such as convolution neural network (CNN), long short term memory (LSTM), and Bidirectional long short term memory (BiLSTM), the BILSTM yielded the best F1-score, among the transfer learning models RoBERTa outperforms the others , indicating the effectiveness of the transformers-based approaches in this context. The researchers mentioned their plan for future work to utilize data augmentation techniques and apply other transfer learning models such as DistilBERT.

2.6.3 Detecting Depression and Suicidal Tendencies in Social Media Using Deep Learning and Feature Selection

Another work by Ismail baydili, et al, [41].Focused on detection of depression and suicidal tendencies in social media data with feature selection The researchers used six various datasets as follows: The Suicidal Ideation Detection (SID) dataset, designed to classify suicidal intent, the Reddit dataset for detecting depression symptoms the Mental Health Corpus dataset, which is divided into two categories toxic and non-toxic comments, the Twitter US Airline Sentiment dataset, which classifies tweets as positive, negative, or neutral; the Suicide and Depression Detection dataset; and finally, the Sentiment140 dataset. They suggested a machine learning-based technique combined

with advanced feature selection methods. The combination of Cumulative Weight-based Iterative Neighborhood Component Analysis (CWINCA) as a feature selector and Support Vector Machines (SVMs) as a classifier performed well on diverse datasets, showing its efficiency in detecting risk factors for mental health disorders. The authors also discussed some limitations of their work: its dependence on English-language data which restricts generalization, class imbalance which lowers model performance, and computational limitations . They lastly state that their future objective is to broaden the dataset to include multiple languages.

2.6.4 Depression Detection Using Machine Learning Algorithms

Another study was carried out by Shumaila Aleem,et al, [42], focusing on the application of machine learning techniques for depression diagnosis. The authors reviewed multiple datasets, including social media posts (e.g., **RSDD** with 9000 depressed and 107,000 control users), EEG signals, and standard clinical assessments like PHQ-9 and BDI. They explored a variety of models such as SVM, Random Forest, CNN, LSTM, and hybrid architectures like AiME. Among the top-performing models ,**1D-CNN** on EEG data, and a multikernel SVM . The work highlighted limitations such as small, imbalanced datasets. For future work, the authors recommend the use of larger, diverse datasets and multimodal deep learning approaches to improve clinical relevance.

2.6.5 A Comprehensive Study of Machine Learning Approaches for Predicting Depression

Continue with the mental health disorder Md. Sabab Zulfiker and his colleagues [43], explored multiple approaches for early detection of depression using machine learning classifier based on socio-demographic and psychological features. utilized survey-based dataset collected from 604 Bangladeshi citizens via questionnaire . Various classifiers were evaluated, including KNN (K-nearest-neighbors) ,AdaBoost,XGBoost, Bagging and weighted voting ensemble. AdaBoost achieved the highest F1-score using SelectKBest features selection technique. However, the dataset is limited by regional and temporal constraints that affect its generalizability.

2.6.6 Systematic Review of ML-Based Depression Diagnosis Using Electronic Health Records

Lastly, a systematic review conducted by David Nickson,et al, [44], investigated the application of machine learning algorithms to Electronic Health Records (EHRs) for depression prediction and

diagnosis. The authors analyzed **19 studies** from both primary and secondary care settings, some involving up to **3 million patient visits**. The models employed ranged from logistic regression and support vector machines to random forests and deep learning techniques.The average performance was around **0.8**. Interestingly, traditional regression models performed nearly as well as advanced ML methods. Key limitations included inconsistent definitions of depression, and the dataset is biased toward participants from Western countries. The authors recommended future research to enhance generalizability and diversify dataset sources.

2.7 Conclusion

In short, AI, deep learning, and NLP change how machines understand and interact with human Languages. Deep learning has significantly enhanced NLP by enabling models to learn complex linguistic patterns.These techniques are building intelligent tools that promote human well-being across many areas, especially the field of mental health.

Chapter 3

Recurrent Neural Networks for Mental Health Disorders

Recurrent Neural Networks (RNNs) have been proposed for mental health analysis as simple models that retain knowledge of input sequences. vanilla RNNs face challenges in memorizing long sequence inputs, making them difficult to train effectively in this context.

To overcome these limitations,we explore a more sophisticated recurrent neural network for mental health analysis. We propose deep learning based approaches to enhance the performance of mental health detection systems.

3.1 Recurrent Neural Networks

In mental health classification, the observation sequence may depend on multiple inputs over long historical dependencies. This creates the need for models and neural networks that can map from the entire history of inputs to predict each output, while also allowing recurrent connections. One way to meet these requirements is by using Recurrent Neural Networks (RNNs), which estimate output probabilities based on both current and past inputs, supporting cyclic connections with a sufficient number of hidden units [45].

Recurrent networks are highly valuable for mental health classification. RNNs, as a flexible family of artificial neural network (ANN) architectures, can leverage sequential information by performing the same operation on each element of a sequence where the output at a given time step depends on previous time steps, thus capturing long-distance dependencies. The key advantage of RNNs is their recurrent memory, which enables them to store and utilize previous inputs in the internal state to influence future outputs.

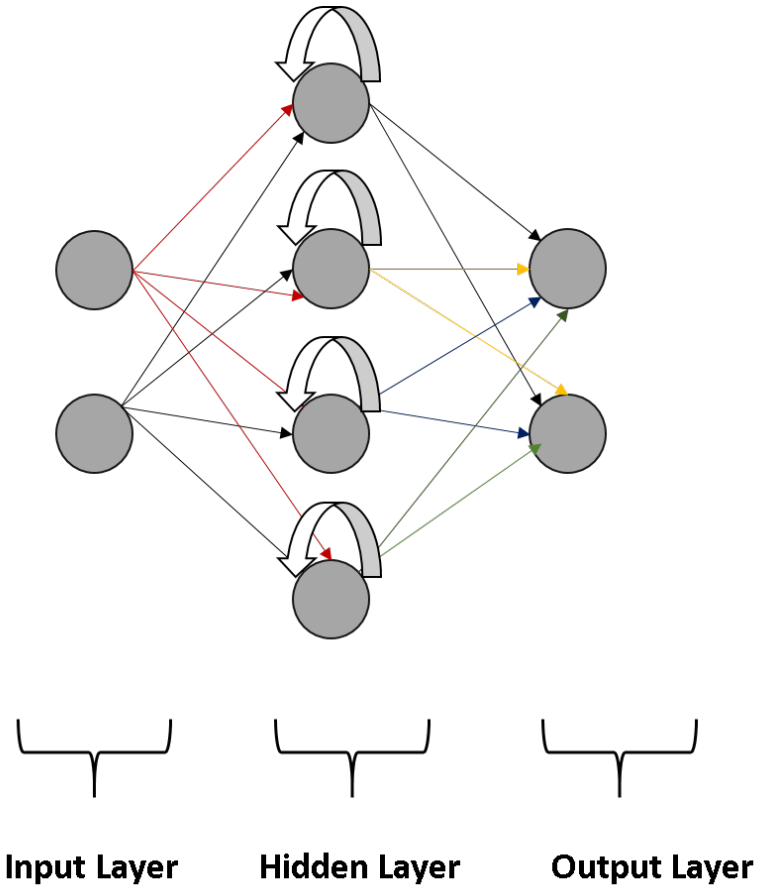


Figure 3.1: General structure of simple RNNs.

An RNN is considered a neural network designed specifically for processing sequences of symbols $(x_1, x_2, \ldots, x_t)$. Most recurrent networks can process sequences of variable length and are more effective for long sequences than standard feed-forward neural networks. Several RNN variants have been proposed, such as Elman networks [46], Jordan networks [47], time-delay neural networks [48], and echo state networks [49].

The structure of widely used RNN models for sequence-based problems like mental health classification typically consists of an input layer, a hidden recurrent layer, and an output layer, as illustrated in Figure 3.1.

A useful way to visualize RNNs is by 'unfolding' the cyclic connections of the network over the input sequence. In Figure 3.2, Section A represents a folded state of RNNs with its corresponding unfolded version in Section B obtained by unrolling the network structure for the complete input

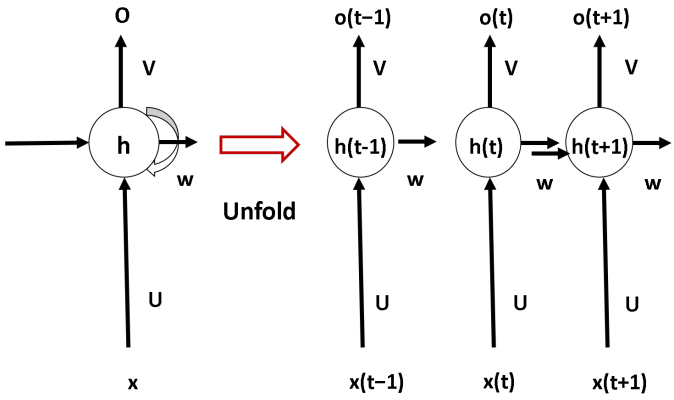


Figure 3.2: General structure of a simple RNN unfolded for three time steps.

sequence, at different and discrete times which in this example contains three-layer neural networks and can be referred to deep neural network because it has more than 1 hidden layer. Note that the unfolded graph, unlike the folded graph, contains no cycles.

- **'U'** represents the weights between the input **'x'** and the hidden state **'h'**.
- **'W'** represents the weights between the hidden states **'h'**.
- **'V'** represents the weights between the hidden states **'h'** and the output **'O'**.

Each node represents a layer of network units at a single time-step. The formulas that govern the computation happening in an RNN are as follows:

- x_t: Input at time step t.
- h_t: Hidden state at time step t. It is the "memory" of the network.h_t is calculated based on the previous hidden state and the input at the current step:

$$h_t = f(Ux_t + Wh_{t-1})$$

  The function f usually is a nonlinearity such as tanh or ReLU. ht−1 is required to calculate the first hidden state, and is typically initialized to zeros.

- o_t: output at the time step t.

3.2 Long Short Term Memory Networks

In this section, we give a detailed description of LSTM networks. We also describe BiLSTM networks which have great influence on the mental health diagnosis as many sequence labeling tasks. The cyclic mechanism enables RNNs to remember inputs at different time steps. They are, therefore, a very good choice for sequence learning. However, because of their difficult training, where gradient descent-based algorithms generally fail to converge or take too much time, or because of the exploding/vanishing gradient problem, which implies that the gradients, during the training, either become very large or very small, their applications in practice were quite limited until the late 1990s. There have been many proposed approaches to diminish the drawbacks when training RNNs, including GRU networks .Among all, LSTM discovered by Hochreiter and Schmidhuber [50] and later refined by Gers [51], appears to be one of the most extensively adopted solutions to the vanishing gradient problem and learn dependencies ranging over arbitrarily long time intervals that have been successfully adopted and used for many sequence modeling tasks.

LSTM networks have a powerful and expressive architecture that has become the most popular variant of RNN to handle sequential data and have been successfully applied to a range of sequence tagging problems such as POS tagging [52] [53], NER [54] [55], sentiment analysis [56] [57] [58], speech recognition [59]. Hochreiter and Schmidhuber [50] proposed to change the basic unit of RNN, which is a simple neuron with a computer memory-like cell, called "LSTM cell". LSTM networks have been made in a specific way. They are the same as RNNs expect the hidden layers were replaced by memory blocks [60],

that have made a difference in their capability to learn long-term dependencies.while RNNs contain cyclic connections in their hidden states, LSTMs still have the recursive connection of RNN with memory cells. The memory blocks store the state over time and have been shown to be better at finding and exploiting long-range dependencies in the data. Hochreiter and Schmidhuber [50] introduce a similar term to that proposed by Cho et al [61]. using gates to prevent limited memory in RNNs. As shown in Figure 3.3. A memory block contains one or more memory cells: LSTM has the ability to add or to remove information from the memory cell that is controlled and protected by gates. A memory block is composed mainly of three gates:

- Input gate.
- Forget gate.
- Output gate.

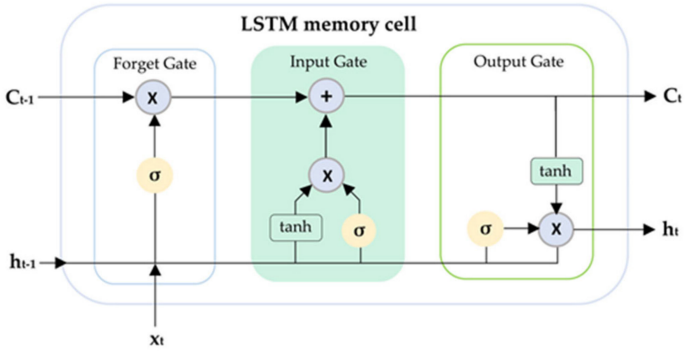


Figure 3.3: Long Short-Term Memory network architecture. [1]

- **Input**: The LSTM unit takes the current input vector at time step t denoted by x_t and the hidden state of the previous time step denoted by h_{t-1}. The sum of the weighted input and hidden state is passed through an activation function, resulting in x_t:

$$x_t = \sigma\big(W_x \cdot [h_{t-1}, x_t] + b_x\big) \tag{3-1}$$

- **Input gate**: The input gate decides which values will be updated and what information to store in the cell. The input gate reads x_t and h_{t-1}, computes the weighted sum, and applies sigmoid activation:

$$i_t = \sigma\big(W_i \cdot [h_{t-1}, x_t] + b_i\big) \tag{3-2}$$

- **Forget gate**: The forget gate is the mechanism through which an LSTM learns to reset the memory contents when they become old and no longer relevant. This may happen, for example, when the network starts processing a new sequence. To remember or throw away information from the cell state, the forget gate reads x_t and h_{t-1} as inputs and applies a sigmoid activation function to the summed weighted inputs:

$$f_t = \sigma\big(W_f \cdot [h_{t-1}, x_t] + b_f\big) \tag{3-3}$$

- **Memory cell**: The current cell state C_t is computed by forgetting irrelevant information from the previous time step and accepting relevant information from the current input. The result, f_t, is multiplied by the previous cell state C_{t-1} to forget memory contents no longer needed, and summed with the multiplication of the input gate i_t and the candidate cell state $\tilde{C}_t$:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{3-4}$$

where

$$\tilde{C}_t = \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \tag{3-5}$$

- **Output gate**: The output gate decides what parts of the cell state to output from the memory cell. It takes the weighted sum of x_t and h_{t-1} and applies sigmoid activation to control what information flows out of the LSTM unit:

$$o_t = \sigma\big(W_o \cdot [h_{t-1}, x_t] + b_o\big) \tag{3-6}$$

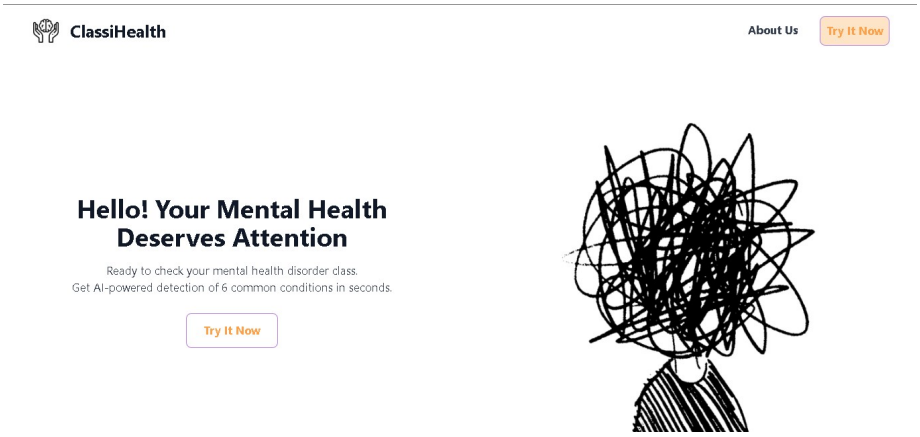
- **Output**: The output of the LSTM unit, h_t, is computed by passing the cell state C_t through a tanh and multiplying it with the output gate O_t:

$$h_t = o_t * \tanh(C_t) \tag{3-7}$$

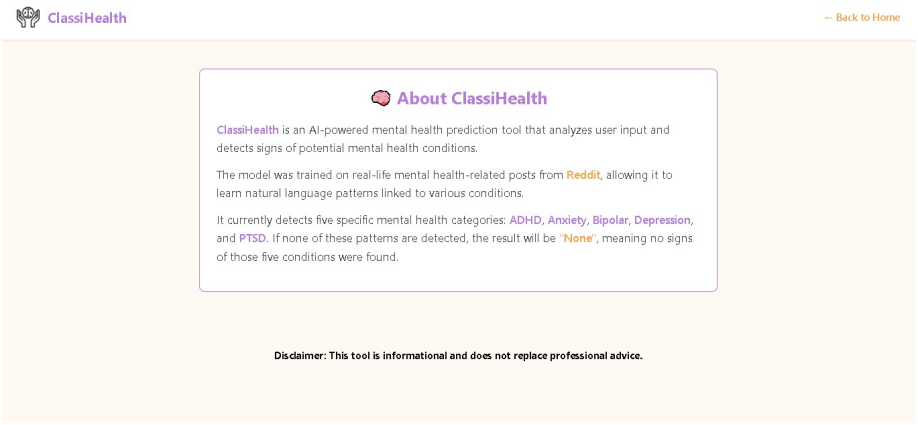
The parameters of the LSTM model are the weight matrices W and bias vectors b.

3.3 Our proposed model for mental health disorders

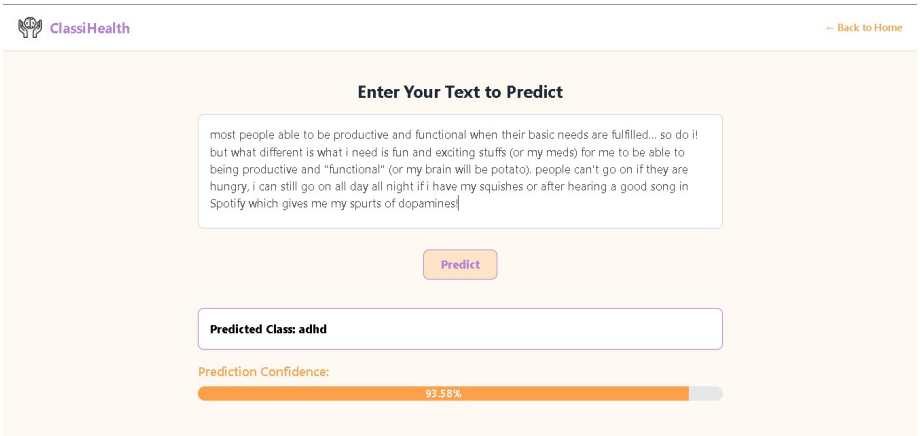
We named our model **Classihealth**, it is an Ai-powered tool designed to help in the prediction of some common mental health disorders including adhd, anxiety, bipolar, ptsd and stress, by analyzing user input (text) using NLP techniques and deep learning models, As demonstrated in Figure 3.4.



(a) Homepage



(b) About us page



(c) Test page

Figure 3.4: Main pages of our tool

Our model consists of three main components:

- input layer
- three different Neural Networks.
- shared output layer.

3.3.1 Input Layer

The input layer varies depending on the model:

- For the first model we used GloVe embeddings
- The second model is based on DistilBERT, which uses contextual embeddings directly from the transformer.
- The third model used RoBERTa embeddings , known for its deep contextual understanding.

3.3.2 Neural Networks

We utilized a combination of deep learning and transformer-based models to classify mental disorders: Backward+BiLSTM (Bidirectional Long Short-Term Memory), DistilBERT, and RoBERTa. The Backward+BiLSTM model was used to capture contextual information from both directions in the text, while DistilBERT and RoBERTa served as transformer-based pretrained models for deeper contextual understanding.

3.3.2.1 Backward + BiLSTM

BiLSTM (Bidirectional Long Short-Term Memory) is a type of recurrent neural network that processes text sequences in both forward and backward directions. This allows it to capture context from both past and future words, making it more effective than standard LSTMs for understanding the full meaning of a sentence, especially in tasks like mental health classification. In our work, we integrated an additional backward LSTM layer into the architecture to enhance the model ability to recognize patterns when processing text in reverse. This adjustment was intended to help the model better capture linguistic patterns often found in mental health-related texts. Overall, this hybrid setup allowed for a deeper understanding of context and led to improved classification performance. This configuration tends to perform better at capturing contextual information, which is valuable in fields like mental health, supported by the findings of [62].

3.3.2.2 Backward + BiLSTM Model Architecture

The backward + BiLSTM architecture is illustrated in Figure 3.5.

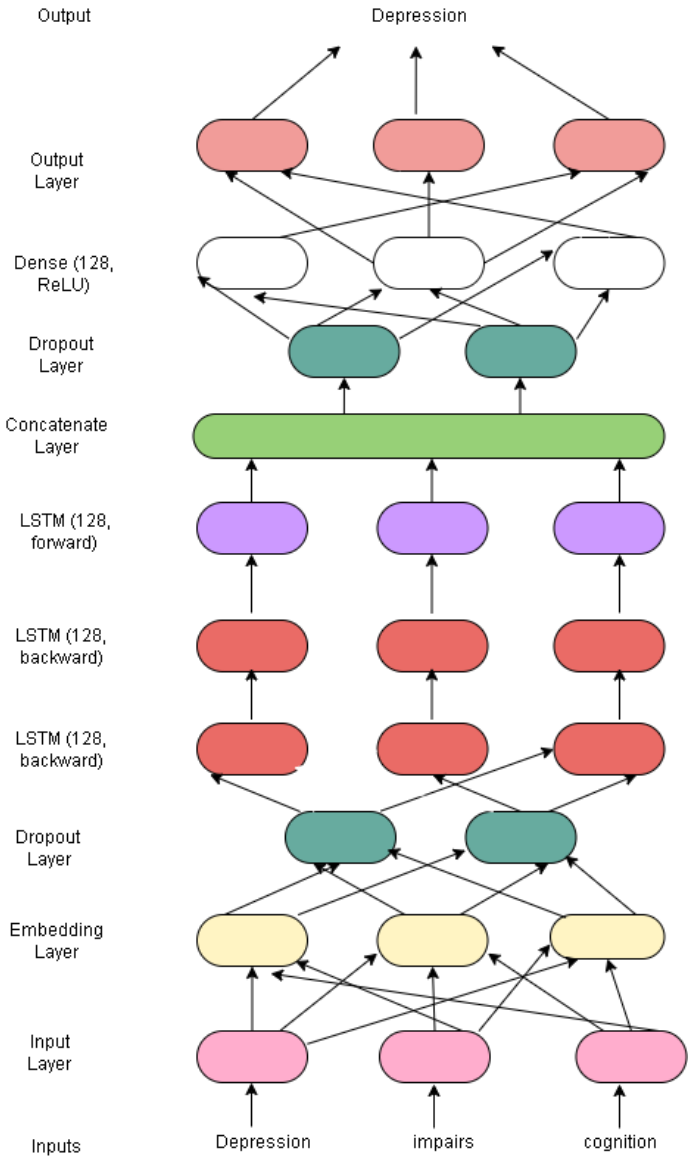


Figure 3.5: Backward + BiLstm Architecture

We began by incorporating pre-trained GloVe embeddings with 50 dimensions to enrich the text

input.
A vocabulary was built from the combined datasets (train, validation, and test), and only words present in the GloVe file (containing 400,000 word vectors) were retained.
Special tokens such as <PAD> (index 0) and <UNK> (index 1) were added to handle padding and out-of-vocabulary words.
Tokens not found in the vocabulary were replaced with <UNK>.

Next, we implemented a backward layer plus a Bidirectional LSTM (BiLSTM) model.
The architecture begins with an embedding layer (50 dim) initialized with the GloVe matrix.
We added a dropout of 0.3 after the embedding to reduce overfitting.
Then, we applied two backward LSTM layers and one forward LSTM, each with 128 units, to capture bidirectional context.
Their outputs were concatenated and passed through a 0.5 dropout layer, followed by a dense layer with 128 units with ReLU activation.
Finally, a softmax output layer with 6 units was used for classification.

The model was compiled using the Adam optimizer and categorical cross-entropy loss function. We trained the model for 25 epochs with batch size of 32, using early stopping (patience = 5).

3.3.2.3 DistilBERT

DistilBERT is a faster and lighter variant of BERT (Bidirectional Encoder Representations from Transformers). It retains 97% of BERT performance while reducing the model size by 40% and running 60% faster. Distilbert is trained through knowledge distillation of the BERT base model [63]. it can be used in NLP tasks such as text classification and question answering where efficiency and speed are important.

3.3.2.4 DistilBERT-based Model Architecture

Distilbert architecture is illustrated in Figure 4.6.

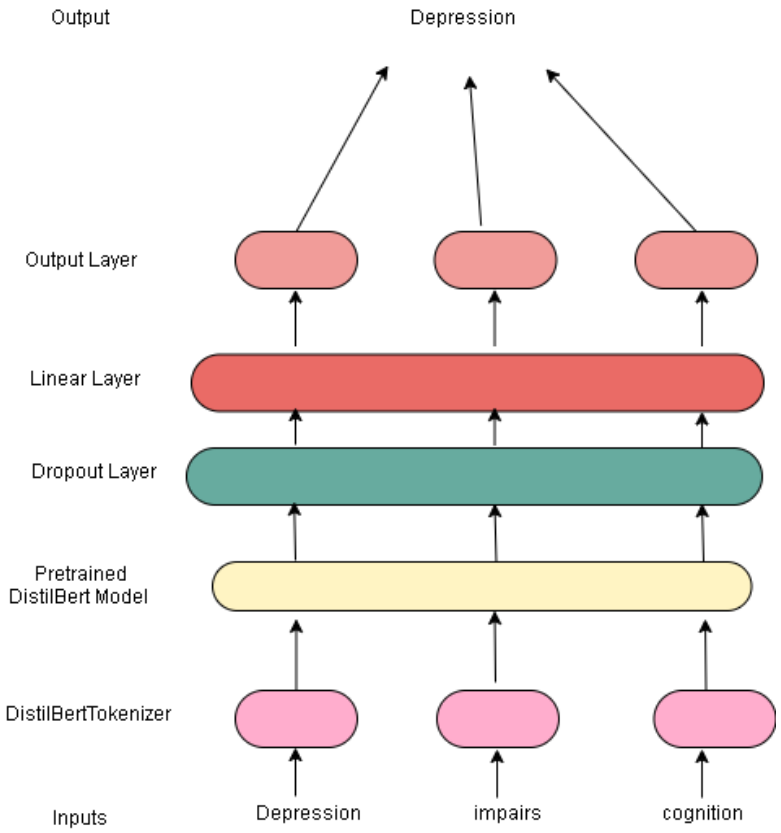


Figure 3.6: DistilBERT Pretrained model architecture

We fine-tuned a DistilBERT model (distilbert-base-uncased) for 6 epochs using PyTorch.
We utilized the DistilBERT tokenizer to tokenize the text data with a max length of 500 tokens, using padding and truncation.
We added a Dropout layer (0.2).
We also used the Adam optimizer with a learning rate of 2e-5, weight decay of 0.01, and cross-entropy loss function.

3.3.2.5 RoBERTa

RoBERTa (Robustly Optimized BERT Approach) is an improved variant of BERT, enhances BERT's performance by training longer, on more data, and removing the next-sentence prediction task. This leads to better understanding of language context, making RoBERTa more accurate for

various NLP tasks like text classification and sentiment analysis [64].

3.3.2.6 RoBERTa-based Model Architecture

RoBERTa architecture is illustrated in Figure 4.7.

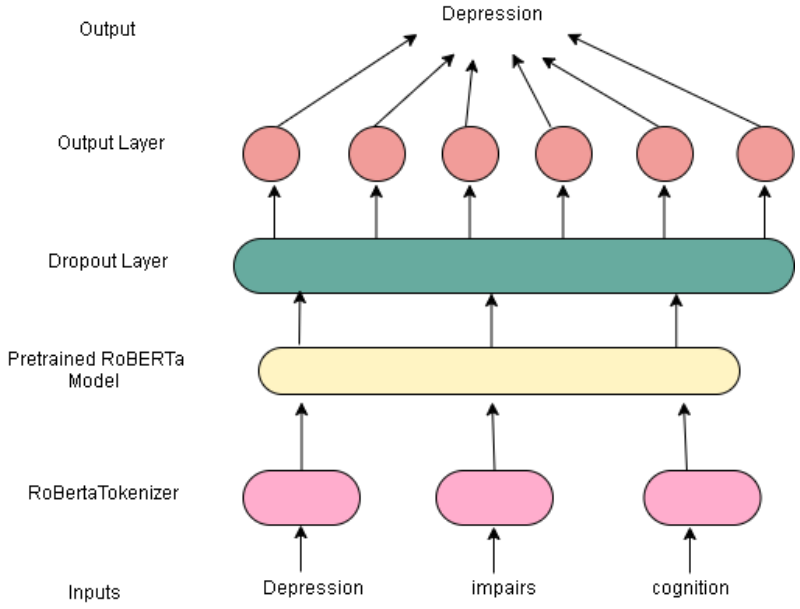


Figure 3.7: RoBerta Pretrained model architecture

We fine-tuned a RoBERTa base model for 10 epochs using PyTorch.
The text inputs were tokenized by the RoBERTa tokenizer with a max length of 500 tokens, using padding and truncation.
we used a Batch size of 32,the AdamW optimizer (learning rate = 1e-5) and cross-entropy loss function.

3.3.3 Output Layer

In all the three models, we used softmax activation function in the output layer. This layer transforms the model final output into a set of six probability values, each represent the likelihood that a given input belongs to one of the six mental health classes. The class with the highest probability is selected as the model prediction.

3.4 Experiment settings

3.5 Dataset

Murarka et al. [39] created multi-class dataset for mental health detection from Reddit social media network, by using the Reddit API.
The dataset includes 16703 posts , split into training, validation and test sets,only 5 mental illnesses were selected, based on the data sufficiency :Bipolar, ADHD, Anxiety, Depression, and PTSD.
To protect users privacy, usersnames and URLs were removed.

Class	**Train**	**Validation**	**Test**
ADHD	2,465	248	248
Anxiety	2,422	248	248
Bipolar	2,407	248	248
Depression	2,450	248	248
PTSD	2,001	248	248
None	1,982	248	248
Total	13,727	1,488	1,488

Table 3.1: Dataset distribution across training, Validation, and test sets for each class

3.6 Data Preprocessing

- We start by checking the dataset for missing and duplicated entries to identify what needs cleaning.
- Next, we prepare a dictionary to expand common English contractions (like “don’t” to “do not”) and apply it to make the text clearer.
- We then detect the language of each entry and keep only English texts.
- To remove noisy data, we filter out words that are too long, have excessive repeated characters, or lack vowels.
- We continue by preprocessing the text: we lowercase it, remove mentions, hashtags, digits, and special characters, clean extra spaces, and tokenize the text.
- After that, we filter out standard stopwords and some frequent non-informative words.
- Then, we split the token lists into chunks of up to 500 tokens for efficient processing and to avoid excessively long posts that could introduce padding noise.

3.6.1 Hyper-parameters and Training

To achieve better performance for each model, we fine-tuned the hyper-parameters accordingly:

3.6.2 Word Embedding

For word embedding, we chose glove (Global Vectors for Word Representation):

3.6.3 GloVe

The GloVe model is an unsupervised algorithm used to represent words as dense vectors by capturing global word co-occurrence statistics from a corpus [65]. Unlike Word2Vec, which relies on local context windows, GloVe constructs a global word-word co-occurrence matrix and factorizes it to generate word embeddings. This allows the model to preserve both syntactic and semantic relationships between words in a more comprehensive manner.

In the context of mental health diagnosis, several studies have demonstrated that integrating GloVe embeddings into deep learning models improves performance. For example, Dey et al [66] conducted a comparative study using LSTM networks with and without GloVe embeddings on mental health chat data. The results showed that the model using GloVe significantly outperformed the one without, achieving higher classification accuracy and faster convergence. These findings support the integration of GloVe to enhance the linguistic understanding of mental health-related language patterns.
In our case, we used glove over word2vec because word2vec has shown almost 10% of out-of-vocabulary (oov) word.

We also, experimented with GloVe embeddings of size 50, 100, 200 and 300, comparing how each dimension influenced the model's understanding of textual data.

3.7 Backward + Bilstm hyper-parameters tuning

3.7.1 Dropout

Dropout is a regularization technique used to prevent overfitting by randomly setting a fraction of the input units to zero during training [67]. In our experiments, we tested dropout values in the range from 0.2 to 0.5, progressively increasing the regularization strength to observe its impact on model generalization.

3.7.2 Hidden LSTM Units

The number of hidden units controls the model's capacity to learn complex patterns. We tested hidden layer sizes from 64 to 256 units, aiming to find a balance between performance and efficiency.

3.7.3 BiLSTM Architecture

BiLSTM models process sequences in both forward and backward directions, enhancing context awareness. To further enrich this, we tested two architectural setups:

- Standard BiLSTM.

- Backward LSTM layer + BiLSTM.to reinforce reverse-sequence learning.

3.7.4 Activation Function

Since this is a multi-class classification task, we used the Softmax function, as it is the most commonly adopted choice for such tasks.

3.7.5 Optimizer

we tested only the Adaptive Moment Estimation (ADAM) Optimizer, since it combines the momentum and RMSprop techniques to provide a more balanced and efficient optimization process [68].

3.7.6 Loss function

We used categorical-crossentropy as our loss function because our task involves multi-class classification. Since this function expects the target labels in one-hot encoded format, we converted our integer class labels accordingly.

Hyperparameter	Values Tested
Dropout Rate	0.2, 0.3, 0.4, 0.5
Hidden Layer Size	64, 128, 256
BiLSTM Configuration	BiLSTM only, backward layer + BiLSTM
GloVe Embedding Dimension	50, 100, 200, 300
Batch Size	8, 16, 32
Optimizer	Adam
Activation Function	Softmax
Loss Function	Categorical-crossentropy

Table 3.2: Summary Table of Hyperparameter Values Tested

After those experiments and tests we choosed the best configuration as follow :

Hyperparameter	Chosen Value
Dropout Rate 1	0.3
Dropout Rate 2	0.5
Hidden LSTM Units	128
Embedding Dimension (GloVe)	50
Optimizer	Adam
Loss Function	Categorical Crossentropy
Activation Function	Softmax

Table 3.3: Best Hyperparameter Configuration

3.8 DistilBERT and RoBERTa hyper-parameters tuning

3.8.1 Dropout

To reduce overfitting, we introduced dropout [67] regularization layers with rates ranging from 0.2 to 0.5.

3.8.2 Learning Rate

We experimented with different learning rates to find the most stable and effective value for fine-tuning DistilBERT and RoBERTa. The tested values ranged from 1e-5 to 5e-5.

3.8.3 Epochs

We evaluated the models over multiple training durations, with epochs ranging from 3 to 10.

3.8.4 Batch Size

We experimented with different batch sizes from 8 to 32 to balance training stability and GPU memory limits.

3.8.5 Activation Function

For both models, the Softmax activation function was used in the output layer to generate probability distributions over the six target classes.

3.8.6 Optimizer

We used the AdamW optimizer, as it is suitable for transformer architectures and helps handle weight decay during training [69].

3.8.7 Loss Function

We applied the standard Cross Entropy Loss, compatible with multi-class classification tasks like ours.

Hyperparameter	**Tested Values**
Learning Rate	1e-5, 2e-5, 3e-5, 4e-5, 5e-5
Epochs	3, 4, 5, 6, 7, 8, 9, 10
Dropout Rate	0.2, 0.3, 0.4, 0.5
Batch Size	8, 16, 32
Optimizer	AdamW
Loss Function	Cross-Entropy

Table 3.4: Tested Hyperparameter for DistilBERT and RoBERTa

After experimenting with different combinations of hyperparametere values for each model, the table below presents the best-performing configuration for both DistilBERT and RoBERTa.

Hyperparameter	**DistilBERT**	**RoBERTa**
Learning Rate	2e-5	1e-5
Epochs	6	10
Dropout Rate	0.2	0.1
Batch Size	16	32
Optimizer	AdamW	AdamW
Loss Function	Cross-Entropy	Cross-Entropy

Table 3.5: Best Hyperparameter Configuration for DistilBERT and RoBERTa

3.9 Conclusion

In this chapter, we presented our contributions to mental health analysis using deep learning and NLP techniques. After identifying the limitations of traditional RNNs, we explored the use of BiLSTM networks to better capture sequencial dependencies in user generated text. Additionally, we also leveraged the strengths of transformer based models, DistilBERT and RoBERTa which provide deep contextual understanding through attention mechanisms and pretraining on large-scale corpora.

Chapter 4

Results and Analysis

This chapter presents the results of experiments conducted using three models: BiLSTM, DistilBERT, and RoBERTa.The models are evaluated using accuracy, precision, recall, and F1-score.Then a comparative analysis was provided, in order to assess the effectiveness of these models in detecting mental health disorders from textual data.

4.1 Tools for Implementation

We implemented our models using the **Python** programming language, which is not only simple but also has a powerful ecosystem for machine and deep learning. We used libraries such as **Keras** and **PyTorch** for building and training neural networks, **Pandas** for handling structured data, and **NumPy** for numerical operations. We also utilized **Scikit-learn** to perform evaluation metrics such as precision, recall, and F1-score. For plotting confusion matrices, we employed **Matplotlib**. For natural language processing tasks, we integrated tools such as **NLTK**.

To build a user-friendly interface, we used **HTML**, **CSS**, and **JavaScript**. This combination of tools enabled an efficient development process and a functional final product.

All implementations were carried out in the **Kaggle environment**, utilizing its **GPU support** to accelerate model training.

4.2 Evaluation Metrics

In machine learning and deep learning, various evaluation metrics are used. Depending on the our task, we selected the appropriate metrics as follows:

4.2.1 Accuracy

Accuracy: measures how correct your model predictions are by calculating the proportion of correct results among the total number of results. The formula for Accuracy is:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

4.2.2 Precision

Precision: refers to the proportion of correct positive predictions (True Positives) out of all the positive predictions made by the model (True Positives + False Positives). It is a measure of the accuracy of the positive predictions. The formula for Precision is:

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

4.2.3 Recall

Recall: is also known as Sensitivity or True Positive Rate where we measures the proportion of actual positive instances that were correctly identified by the model. It is the ratio of True Positives to the total actual positives (True Positives + False Negatives). The formula for Recall is:

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

4.2.4 F1-score

F1-Score: is the combination of both precision and recall using the harmonic mean:

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

4.2.5 Confusion Matrix

is a simple table used to measure how well a classification model is performing. It compares the predictions made by the model with the actual results and shows where the model was right or wrong. This helps you understand where the model is making mistakes so you can improve it. It breaks down the predictions into four categories:

- TP: True Positive
- TN: True Negative

- FP: False Positive
- FN: False Negative

	Predicted Positive	Predicted Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

4.3 Performance Analysis

4.3.1 Backward + BiLSTM Model Performance

Table 4.1 reports the performance measures over each class using the Backward + BiLSTM model and table 4.2 shows the overall model performance.

Class	**Precision**	**Recall**	**F1-score**	**Support**
ADHD	71%	81%	76%	252
ANXIETY	77%	68%	72%	254
BIPOLAR	84%	64%	73%	253
DEPRESSION	55%	80%	65%	249
PTSD	81%	70%	75%	255
NONE	97%	87%	92%	264

Table 4.1: Backward + BiLSTM Per-Class Evaluation Metrics

Model	**Precision**	**Recall**	**F1-score**	**Accuracy**
BiLSTM	77.49%	75.12%	75.47%	75.18%

Table 4.2: Performance Metrics of Backward + BiLSTM Model

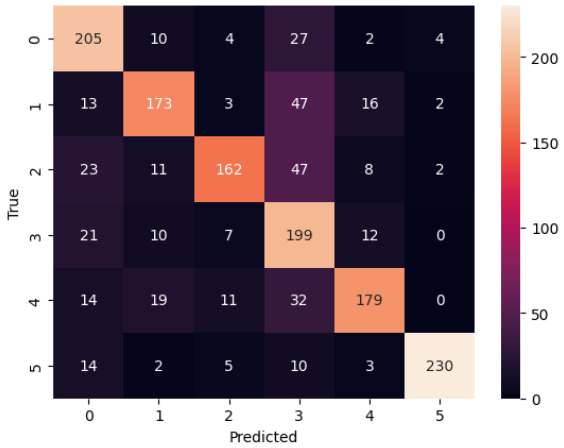


Figure 4.1: Confusion Matrix of Backward + BiLSTM Model

4.3.2 DistilBERT Model Performance

Table 4.3 reports the performance measures over each class using the Backward + BiLSTM model and table 4.4 shows the overall model performance.

Class	Precision	Recall	F1-score	Support
ADHD	81%	84%	82%	252
ANXIETY	74%	81%	77%	254
BIPOLAR	81%	75%	78%	253
DEPRESSION	74%	72%	73%	249
PTSD	85%	82%	83%	255
NONE	95%	96%	96%	264

Table 4.3: DistilBERT Per-Class Evaluation Metrics

Model	Precision	Recall	F1-score	Accuracy
DistilBERT	82%	82.76%	82.66%	82.66%

Table 4.4: Performance Metrics of DistilBERT Model

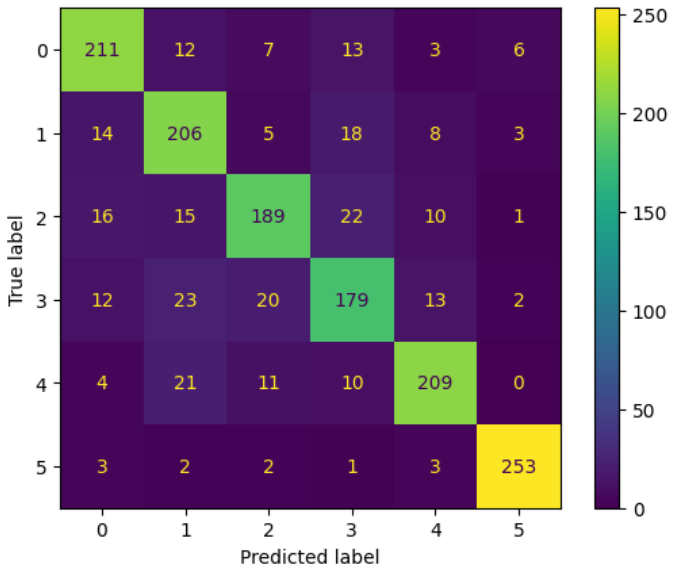


Figure 4.2: Confusion Matrix of DistilBERT Model

4.3.3 Performance Metrics of RoBERTa Model

Table 4.5 reports the performance measures over each class using the Backward + BiLSTM model and table 4.6 shows the overall model performance.

Class	Precision	Recall	F1-score	Support
ADHD	85%	88%	86%	252
ANXIETY	78%	82%	80%	254
BIPOLAR	86%	80%	83%	253
DEPRESSION	76%	82%	79%	249
PTSD	88%	82%	85%	255
NONE	99%	97%	98%	264

Table 4.5: RoBERTa Per-Class Evaluation Metrics

Model	Precision	Recall	F1-score	Accuracy
RoBERTa	86.00%	85.00%	85.00%	85.33%

Table 4.6: Performance Metrics of RoBERTa Model

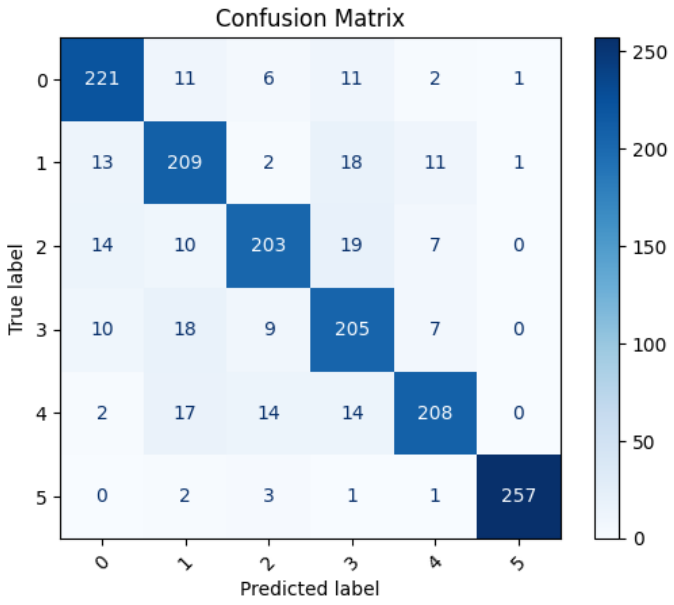


Figure 4.3: Confusion Matrix of RoBERTa Model

4.4 Performance Comparison of Our Models

The analysis highlights that RoBERTa outperforms BiLSTM and DistilBERT models, by achieving the highest accuracy of 85.33% with balanced recall and precision scores of approximately 85.00%, demonstrating its capability to effectively classify mental disorders classes while reducing the occurrence of false positives and false negatives.

DistilBERT achieved a high accuracy of 81.66% with good precision and recall value particularly when considering its quicker inference times and lower computing cost compared to RoBERTa.

In contrast, the BiLSTM model, performs less effectively with an accuracy of approximately 75.18% ,with lower precision and recall scores, indicates that it is less capable of capturing the complex linguistic patterns within the dataset than transformer-based models.

Based on per-class performance, RoBERTa performs well across all classes, particularly in detecting the 'NONE' and 'ADHD' classes.

In this section, we compare the performance of our models with baseline models reported in previous studies.

We compared the performance of our three deep learning models BiLSTM, DistilBERT, and RoBERTa based on four primary measures: accuracy, precision, recall, and F1-score.

Model	Accuracy	Precision	Recall	F1-score
BiLSTM	75.18%	77.49%	75.12%	75.47%
DistilBERT	81.66%	81.76%	81.66%	81.66%
RoBERTa	**85.33%**	**86.00%**	**85.00%**	**85.00%**

Table 4.7: Performance comparison between our models

The BiLSTM model was a good baseline, with accuracy of 75.18% and F1-score of 75.47%. Although it performed well,it did not perform as well as the transformer-based models,showing its weakness at understanding contextual representation efficiently compared to pretrained language models.

DistilBERT performed significantly better than BiLSTM with an accuracy of 81.66% and an F1-score of 81.66%. This shows that even a smaller transformer model can perform well when it is trained on large scale corpora , offering a good balance between performance and speed. RoBERTa outperformed its predecessor models, achieving the highest results across all metrics,with an accuracy of 85.33%, precision of 86.00%, recall of 85.00%, and an F1-score of 85.00%. These results highlight how well RoBERTa can understand complex linguistics structures.

4.5 Comparative Analysis with Previous Studies

Model	Accuracy	Precision	Recall	F1-score
LSTM [39]	72%	74%	72%	72%
NB [40]	66.49%	72.18%	66.73%	66.73%
RF [40]	70.85%	72.46%	70.50%	70.50%
LR [40]	77.87%	78.24%	77.89%	77.89%
CNN [40]	–	82.84%	82.65%	81.64%
AlBERT [40]	–	80.90%	80.38%	80.45%
Backward + BiLSTM (our)	75.18%	77.49%	75.12%	75.47%
DistilBERT (our)	81.66%	81.76%	81.66%	81.66%
RoBERTa (our)	**85.33%**	**86.00%**	**85.00%**	**85.00%**

Table 4.8: Comparison between our models and models from previous studies.

The results in Table 4.8 demonstrate the performance of our models Backward+BiLSTM, DistilBERT, and RoBERTa compared to previous deep learning models tested on post-level data.

Backward+BiLSTM, while not as strong as the transformer-based models, still outperformed traditional baselines such as Naive Bayes, Random Forest, and earlier LSTM models. This highlights the value of our model in capturing contextual information more effectively than previous models.

The DistilBERT model also performed strongly, as indicated by its balanced precision, recall, and F1-score (all above 81.00%), outperforming both traditional machine learning methods and earlier deep learning models.

Among the three models, RoBERTa achieved the highest scores across all evaluation metrics, with 85.33% accuracy, 86.00% precision, 85.00% recall, and an F1-score of 85.00%. These results surpass those of all previously developed models.

4.6 Conclusion

In conclusion, the analysis confirms that transfer learning models particularly RoBERTa, show better performance in mental health classification tasks, highlighting the relevance of transformer-based models in text classification tasks over traditional machine learning methods.

Conclusion

In this thesis, we investigated and developed different techniques and approaches for Mental health disorders detection. In particular, we carried out on applying deep learning based methods; we have worked on the Mental Health Reddit data for the problem. The major contributions of this thesis are summarized below:

First, a new approach based on LSTM nets is presented. The proposed method is based on BLSTMs. A Backward LSTM is introduced to combine input lookup tables. Then the current de facto standard in sequence labeling tasks: BLSTM based neural net is used and a Softmax activation function is used to predict the final outputs. We tested the efficiency of our proposed method. The experimental results show that the Backward-BLSTM technique is efficient for the task.

After that, an adapted DistilBERT proposed architecture was proposed for the task as an efficient Transformer model that retains most of BERT39;s performance while significantly reducing computational cost. Evaluations on the test set comparing to state-of-the-arts demonstrate the effectiveness of the proposed model.

Then, a RoBERTa based architecture was also proposed as a more robust and accurate Transformer model. The objective of usig RoBERTa is to benefit from its deep contextual understanding to accurately categorize text into one of several predefined classes (outputs).

In summary, all the models described in this thesis are very simple and efficient for automatic mental health disorders detection of English language text. The models have much higher accuracy than the machine learning models and are state-of-the-art. Further work in this area could be done in several directions. First, We aim to improve the models performance by incorporating a wide variety of real and diverse datasets while focusing on more significant features to enhance generalizability and reduce bias. We also plan to collaborate with mental health clinicians to ensure that the model and its predictions hold clinical validity and are suitable for real-world applications. Moreover, We intend to scale the system to support multiple languages and adapt it to various cultural frameworks, enabling broader accessibility in representing mental health conditions.

Bibliography

[1] K. Nifa, A. Boudhar, H. Ouatiki, H. Elyoussfi, B. Bargam, and A. Chehbouni, "Deep learning approach with lstm for daily streamflow prediction in a semi-arid area: A case study of oum er-rbia river basin, morocco," *Water*, vol. 15, p. 262, 01 2023.

[2] P. H. Somerset, "The little book of mental health." https://www.cypsomersethealth.org/images/The_little_book_of_mental_health_-_version_6.pdf, 2020. Accessed: 14 June 2025.

[3] W. H. Organization, "Mental health." https://www.who.int/health-topics/mental-health#tab=tab_2, 2025. Accessed: 22 Mar. 2025.

[4] W. H. Organization, "Mental health stat." https://www.who.int/news-room/fact-sheets/detail/mental-disorders, 2025. Accessed: 22 Mar. 2025.

[5] M. clinic, "Depression." https://www.mayoclinic.org/diseases-conditions/depression/symptoms-causes/syc-20356007, 2007. NIH Publication No. 07-3561, U.S. Department of Health and Human Services.

[6] W. H. Organization, "Anxiety definition." https://www.nhs.uk/every-mind-matters/mental-health-issues/anxiety/, 2025. Accessed: 22 Mar. 2025.

[7] W. H. Organization, "Adhd efinition." https://www.psychiatry.org/patients-families/adhd/what-is-adhd#:~:text=Attention%2Ddeficit%2Fhyperactivity%20disorder%20(ADHD)%20is%20one%20of,in%20the%20moment%20without%20thought)., 2025. Accessed: 22 Mar. 2025.

[8] W. H. Organization, "Ptsd definition." https://www.psychiatry.org/patients-families/ptsd/what-is-ptsd, 2025. Accessed: 22 Mar. 2025.

[9] W. H. Organization, "Bipolar definition." https://www.nimh.nih.gov/health/topics/bipolar-disorder, 2025. Accessed: 22 Mar. 2025.

[10] Mayo Clinic Staff, "Depression (major depressive disorder)." https://www.mayoclinic.org/diseases-conditions/depression/symptoms-causes/syc-20356007, n.d. Accessed: May 2025.

[11] Mayo Clinic Staff, "Generalized anxiety disorder." https://www.mayoclinic.org/diseases-conditions/anxiety/symptoms-causes/syc-20350961, n.d. Accessed: May 2025.

[12] Mayo Clinic Staff, "Adult adhd." https://www.mayoclinic.org/diseases-conditions/adult-adhd/symptoms-causes/syc-20350878, n.d. Accessed: May 2025.

[13] Mayo Clinic Staff, "Post-traumatic stress disorder (ptsd)." https://www.mayoclinic.org/diseases-conditions/post-traumatic-stress-disorder/symptoms-causes/syc-20355967, n.d. Accessed: May 2025.

[14] Mayo Clinic Staff, "Bipolar disorder." https://www.mayoclinic.org/diseases-conditions/bipolar-disorder/symptoms-causes/syc-20355955, n.d. Accessed: May 2025.

[15] Mayo Clinic, "Mental illness: Symptoms and causes." https://www.mayoclinic.org/diseases-conditions/mental-illness/symptoms-causes/syc-20374968, 2022. Accessed: 2025-04-26.

[16] Mayo Clinic, "Mental illness: diagnosis." https://www.mayoclinic.org/diseases-conditions/mental-illness/diagnosis-treatment/drc-20374974. Accessed: 2025-04-27.

[17] Mayo Clinic Staff, "Mental illness – diagnosis and treatment." https://www.mayoclinic.org/diseases-conditions/mental-illness/diagnosis-treatment/drc-20374974, 2025. Accessed: May 27, 2025.

[18] Mayo Clinic Staff, "Mental illness – symptoms and causes." https://www.mayoclinic.org/diseases-conditions/mental-illness/symptoms-causes/syc-20374968, 2025. Accessed: May 27, 2025.

[19] H. Sheikh, C. Prins, and E. Schrijvers, *Mission AI: The New System Technology*. Research for Policy, Cham: Springer, 2023.

[20] R. Dehbozorgi, S. Zangeneh, E. Khooshab, D. H. Nia, H. R. Hanif, P. Samian, M. Yousefi, F. H. Hashemi, M. Vakili, N. Jamalimoghadam, and F. Lohrasebi, "The application of artificial intelligence in the field of mental health: a systematic review," *BMC Psychiatry*, vol. 25, p. 132, Feb. 2025.

[21] J. A. Nichols, H. W. H. Chan, and M. A. B. Baker, "Machine learning: Applications of artificial intelligence to imaging and diagnosis," *Biophysical Reviews*, vol. 11, no. 1, pp. 111–118, 2019.

[22] A. Mathew, A. Arul, and S. Sivakumari, *Deep Learning Techniques: An Overview*, pp. 599–608. 01 2021.

[23] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by backpropagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[24] K. D. Kannan, B. Senthilkumar, N. M. Gnanasekar, *et al.*, "Advancements in machine learning and deep learning for early detection and management of mental health disorder," *arXiv preprint arXiv:2412.06147*, 2024.

[25] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[26] Y. Bengio *et al.*, "Learning deep architectures for ai," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[27] M. Ranzato and M. Szummer, "Semi-supervised learning of compact document representations with deep networks," in *Proceedings of the 25th International Conference on Machine Learning*, pp. 792–799, 2008.

[28] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proceedings of the 25th International Conference on Machine Learning*, pp. 160–167, 2008.

[29] A. Mnih and G. E. Hinton, "A scalable hierarchical distributed language model," in *Advances in Neural Information Processing Systems*, pp. 1081–1088, 2009.

[30] J. Weston, F. Ratle, H. Mobahi, *et al.*, "Deep learning via semi-supervised embedding," in *Neural Networks: Tricks of the Trade*, pp. 639–655, Springer, 2012.

[31] I. Arel, D. C. Rose, and T. P. Karnowski, "Deep machine learning—a new frontier in artificial intelligence research," *IEEE Computational Intelligence Magazine*, vol. 5, no. 4, pp. 13–18, 2010.

[32] D. Yu and L. Deng, "Deep learning and its applications to signal and information processing," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 145–154, 2011.

[33] G. Hinton, L. Deng, D. Yu, *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[34] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[35] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*. 3rd ed., 2025. Online manuscript released January 12, 2025.

[36] A. H. Saffar, T. K. Mann, and B. Ofoghi, "Textual emotion detection in health: Advances and applications," *Journal of Biomedical Informatics*, vol. 137, p. 104258, 2023.

[37] V. Dogra, P. Verma, K. ., P. Chatterjee, J. Shafi, C. Jaeyoung, and M. F. Ijaz, "A complete process of text classification system using state-of-the-art nlp models," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–26, 06 2022.

[38] A. Aqlan, D. M. Bairam, and R. L. Naik, *A Study of Sentiment Analysis: Concepts, Techniques, and Challenges*, pp. 147–162. 01 2019.

[39] A. Murarka, B. Radhakrishnan, and S. Ravichandran, "Classification of mental illnesses on social media using RoBERTa," in *Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis* (E. Holderness, A. Jimeno Yepes, A. Lavelli, A.-L. Minard, J. Pustejovsky, and F. Rinaldi, eds.), (online), pp. 59–68, Association for Computational Linguistics, Apr. 2021.

[40] I. Ameer, N. Bölücü, G. Sidorov, A. Gelbukh, and V. Elangovan, "Mental illness classification on social media texts using deep learning and transfer learning," *Computacion y Sistemas*, vol. 28, pp. 451–464, 06 2024.

[41] I. Baydili, B. Taşcı, and G. Tasci, "Deep learning-based detection of depression and suicidal tendencies in social media data with feature selection," *Behavioral Sciences*, vol. 15, 03 2025.

[42] S. Aleem, N. U. Huda, R. Amin, S. Khalid, S. Alshamrani, and A. Alshehri, "Machine learning algorithms for depression: Diagnosis, insights, and research directions," *Electronics*, vol. 11, p. 1111, 03 2022.

[43] M. Zulfiker, N. Ety, A. A. Biswas, T. Nazneen, and M. S. Uddin, "An in-depth analysis of machine learning approaches to predict depression," *Current Research in Behavioral Sciences*, vol. 2, p. 100044, 05 2021.

[44] D. Nickson, C. Meyer, L. Walasek, and C. Toro, "Prediction and diagnosis of depression using machine learning with electronic health records data: a systematic review," *BMC Medical Informatics and Decision Making*, vol. 23, 11 2023.

[45] B. Hammer, "On the approximation capability of recurrent neural networks," *Neurocomputing*, vol. 31, no. 1-4, pp. 107–123, 2000.

[46] J. L. Elman, "Finding structure in time," *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990.

[47] M. Jordan, "Attractor dynamics and parallelism in a connectionist sequential machine," in *Proc. 8th Annu. Conf. Cognitive Science Society*, pp. 513–546, 1986.

[48] K. J. Lang, A. H. Waibel, and G. E. Hinton, "A time-delay neural network architecture for isolated word recognition," *Neural Networks*, vol. 3, no. 1, pp. 23–43, 1990.

[49] H. Jaeger, "The "echo state" approach to analysing and training recurrent neural networks—with an erratum note," Tech. Rep. 148(34), German National Research Center for Information Technology, GMD, Bonn, Germany, 2001.

[50] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[51] F. Gers, *Long Short-Term Memory in Recurrent Neural Networks*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2001. Unpublished PhD dissertation.

[52] M. Peters, W. Ammar, C. Bhagavatula, *et al.*, "Semi-supervised sequence tagging with bidirectional language models," in *Proc. 55th Annu. Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, pp. 1756–1765, 2017.

[53] B. Plank, A. Søgaard, and Y. Goldberg, "Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss," in *Proc. 54th Annu. Meeting of the Association for Computational Linguistics*, p. 412, 2016.

[54] N. Limsopatham and N. Collier, "Bidirectional lstm for named entity recognition in twitter messages," in *Proc. Workshop on Noisy User-generated Text (WNUT)*, p. 145, 2016.

[55] S. Yan, C. Hardmeier, and J. Nivre, "Multilingual named entity recognition using hybrid neural networks," in *Proc. 6th Swedish Language Technology Conference (SLTC)*, 2016.

[56] D. Tang, B. Qin, X. Feng, *et al.*, "Effective lstms for target-dependent sentiment classification," in *Proc. COLING 2016, 26th Int. Conf. Computational Linguistics: Technical Papers*, pp. 3298–3307, 2016.

[57] Y. Wang, M. Huang, L. Zhao, *et al.*, "Attention-based lstm for aspect-level sentiment classification," in *Proc. 2016 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, pp. 606–615, 2016.

[58] M. Yang, W. Tu, J. Wang, *et al.*, "Attention based lstm for target dependent sentiment classification," in *Proc. AAAI Conf. Artificial Intelligence*, pp. 5013–5014, 2017.

[59] A. Graves, A. r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6645–6649, 2013.

[60] F. Gers, *Long Short-Term Memory in Recurrent Neural Networks*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2001. Unpublished Ph.D. dissertation.

[61] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder–decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734, 2014.

[62] R. Kadari, Y. Zhang, W. Zhang, and T. Liu, "Ccg supertagging with bidirectional long short-term memory networks," *Natural Language Engineering*, vol. 24, no. 1, pp. 77–90, 2018.

[63] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter," 10 2019.

[64] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," 07 2019.

[65] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," vol. 14, pp. 1532–1543, 01 2014.

[66] J. Dey and D. Desai, "Nlp based approach for classification of mental health issues using lstm and glove embeddings," *International Journal of Advanced Research in Science, Communication and Technology*, pp. 347–354, 01 2022.

[67] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 06 2014.

[68] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[69] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.

ملخص

يُطلق على التفاعل مع الحواسيب باستخدام أنظمة الذكاء الاصطناعي من خلال اللغات الطبيعية اسم **معالجة اللغة الطبيعية (NLP)**. وتُعد معالجة اللغة الطبيعية من المجالات التي لها تطبيقات عديدة تُستخدم على نطاق واسع في حياتنا اليومية. يُعتبر **تصنيف التسلسلات** من أقدم المجالات في معالجة اللغة الطبيعية، ويشمل العديد من المهام. وقد أصبحت **تنبؤات الاضطرابات النفسية** من بين المهام الأكثر أهمية في السنوات الأخيرة، حيث تم اقتراح مقاربات فعالة تعتمد على تقنيات التعلم الآلي.

في هذا العمل، ندرس تطبيق التعلم العميق لتصنيف الاضطرابات النفسية اعتمادًا على النصوص المكتوبة من قبل المستخدمين. لقد قمنا باقتراح ثلاث معماريات مختلفة قائمة على التعلم العميق، وهي :**شبكةBILSTM** ، والمحول الفعّال **DistilBERT**، والمحول المُحسَّن بدقة .**RoBERTa** وقد أكدت النتائج التجريبية فعالية التقنيات المقترحة، حيث حقق نموذج **RoBERTa** أفضل أداء في التصنيف بين جميع النماذج المُقترحة

الكلمات المفتاحية :معالجة اللغة الطبيعية، الاضطرابات النفسية، BILSTM، DistilBERT، RoBERTa

Abstract

Interacting with computers with AI systems using natural languages is often referred to as **Natural Language Processing (NLP).** NLP has many applications that are widely used in our daily lives. Sequence labeling is one of the oldest fields in NLP including many tasks. Prediction of the mental health disorders has been one of the most important tasks in recent years where dominant approaches based on machine learning methods have been proposed. In this thesis, we explore the application of deep learning for classifying mental disorders based on user-written text. We experimentally proposed three different deep learning based architectures; including **BILSTM,** the efficient transformer **DistilBERT,** and the robustly optimized transformer **RoBERTa.** The achieved experimental results confirm the effectiveness of our proposed techniques. Among the proposed models**, RoBERTa** achieved the highest classification performance.

Keywords : Natural Language Processing (NLP), Mental Health Disorders, BILSTM, DistilBERT, RoBERTa

Résumé

L'interaction avec les ordinateurs à l'aide de systèmes d'intelligence artificielle utilisant les langues naturelles est souvent appelée **Traitement Automatique du Langage Naturel (TALN)**. Le TALN possède de nombreuses applications largement utilisées dans notre vie quotidienne. L'étiquetage de séquences est l'un des domaines les plus anciens du TALN, englobant plusieurs tâches.

La prédiction des troubles de la santé mentale est devenue l'une des tâches les plus importantes ces dernières années, où des approches dominantes basées sur les méthodes d'apprentissage automatique ont été proposées.

Dans ce mémoire, nous étudions l'application de l'apprentissage profond pour la classification des troubles mentaux à partir de textes rédigés par les utilisateurs.

Nous avons proposé expérimentalement trois architectures différentes basées sur l'apprentissage profond, notamment : **BILSTM**, le transformeur efficace **DistilBERT**, et le transformeur optimisé de manière robuste **RoBERTa**.

Les résultats expérimentaux obtenus confirment l'efficacité des techniques proposées. Parmi les modèles proposés, **RoBERTa** a atteint les meilleures performances en matière de classification.

Mots-clés : Traitement Automatique du Langage Naturel (TALN), Troubles de la santé mentale, BILSTM, DistilBERT, RoBERTa.