FINAL YEAR PROJECT

TEXT CLASSIFICATION FOR MAJOR DEPRESSIVE DISORDER (MDD) SYMPTOMS AND TREATMENTS USING CONVOLUTIONAL NEURAL NETWORKS (CNN)

Presentation Video:https://youtu.be/_m0fpcqcHpA
Demo Video:https://youtu.be/fSKtOvFLNkw

PRESENTED BY
FELICIA CHIN HUI FEN
A20EC0037

SUPERVISED BY
DR SHARIN HAZLIN BINTI HUPSI
DR AHMAD NAJMI BIN AMERHAIDER NUAR

TABLES OF CONTENTS

CHAPTER 1 INTRODUCTION

INTRODUCTION

Major Depressive Disorder (MDD)

- commonly known as depression.
- root cause still unknown.
- earlier diagnosis and treatment for MDD can be challenging.

Text Classification

- a type of natural language processing (NLP).
- helps to assign tags or labels to the textual contents.

Convolutional Neural Networks (CNN)

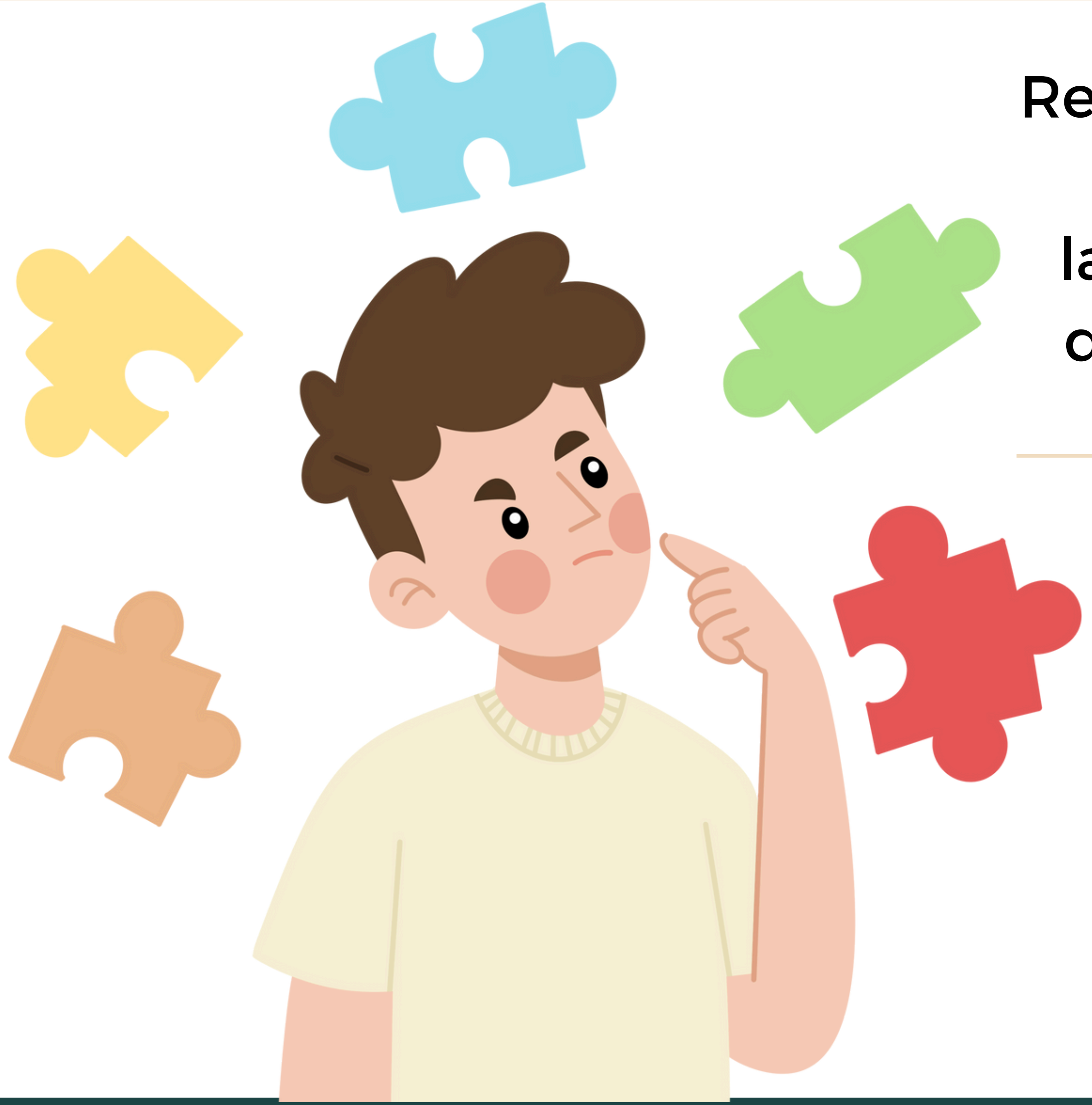
- a type of deep learning method that shows significant results in the text classification.

PROBLEM BACKGROUND

Early detection of symptoms of MDD and seeking treatment is **critical** for anyone who suspects they have this disorder.

MDD treatments are **ineffective** because they require different types of medication, time, planning and are specific to each patient.

Symptomatology is the only method available for pre-diagnosis.



Recent studies have shown the **potential of text classification** applying natural language processing methods to identify depressive symptoms but in using **social media text data.**

Potential to develop a text classification model for MDD symptoms and treatments using CNN based on medical journal datasets.

Back to table of contents

PROBLEM STATEMENT

- MDD is a critical mental disorder that is difficult to diagnose and treat due to its multifactorial nature.
- Identifying MDD symptoms is difficult due to the lack of biological markers.
- Treatment is also unique for each individual, and it takes time and preparation.
- In order to pre-diagnose Major Depressive Disorder (MDD) based on symptoms, an effective and efficient method should be used.
- Seeks to create a text classification model for Major Depressive Disorder (MDD) symptoms and treatments based on text data from medical journals using Convolutional Neural Networks (CNN).

Back to table of contents

RESEARCH OBJECTIVES

Goal

- To develop a classification model for the symptoms and treatments of the Major Depressive Disorder (MDD) using the Convolutional Neural Networks (CNN) based on the text documents in the medical journals.

Objectives

- To identify the related features for Major Depression Disorder (MDD) symptoms and treatments in the medical journals.
- To perform the text classification method for collection of Major Depression Disorder (MDD) symptoms and treatments using Convolutional Neural Networks (CNN) algorithms.
- To evaluate the performance of the machine learning methodologies, Convolutional Neural Networks (CNN) in order to identify the Major Depressive Disorder (MDD) symptoms and treatments.

Back to table of contents

RESEARCH SCOPE

1

The research is conducted using the 5000 medical journals from the NCBI website.

2

The medical journals are chosen by focusing on symptoms and treatments of the MDD.

3

Word2vec and Keras embeddings will be used to create word embeddings for the symptoms and treatments of Major Depression Disorder (MDD).

4

CNN algorithms will be used to classify texts or words among the medical journals.

Back to table of contents

CHAPTER 2
LITERATURE REVIEW

Back to table of contents

SUMMARY OF RELATED WORK

Multi Label Text Classification Using Deep Learning Approaches

- Mohammed et al. (2020) reported that a CNN model ranked second in classifying toxic comment levels and was able to achieve more than 80.00% precision, recall, and F1-score.
- Elnagar et al. (2022) found that CNN had the highest accuracy with a confidence greater than 50%, contributing to 70.34% of the overall performance among the deep learning models.
- Wang et al. (2021) demonstrated that a CNN algorithm achieved more than 50.00% accuracy and a 70.00% macro-F1 score.

Text Classification MDD Symptoms

- Kim et al. (2022) reported that a CNN was able to achieve more than 75.00% accuracy using both sentence classifiers (PHQ-9 and Yes-No classifier).
- CNN model achieved more than 90.00% performance in detecting depression using a Twitter dataset (Amanant et al., 2022; Kour and Gupta, 2022).

Identifying MDD Treatments

- Wang et al. (2019): Identified MDD symptoms and predicted treatment response to discover potential development of antidepressant treatments using biological data.
- Watts et al. (2022): Predicted treatment response for MDD using electronic databases and various machine learning approaches.

Back to table of contents

DISCUSSION

01

Text classification research on MDD symptoms mostly using social media text datasets.

02

Exists of research in text classification on MDD symptoms but without its treatments.

03

Exists of research in identifying MDD treatments using medical journals datasets.

04

Text classification using CNN algorithm is able to obtain good results. It is able to obtained a range of accuracy from 70.00% to 80.00%

05

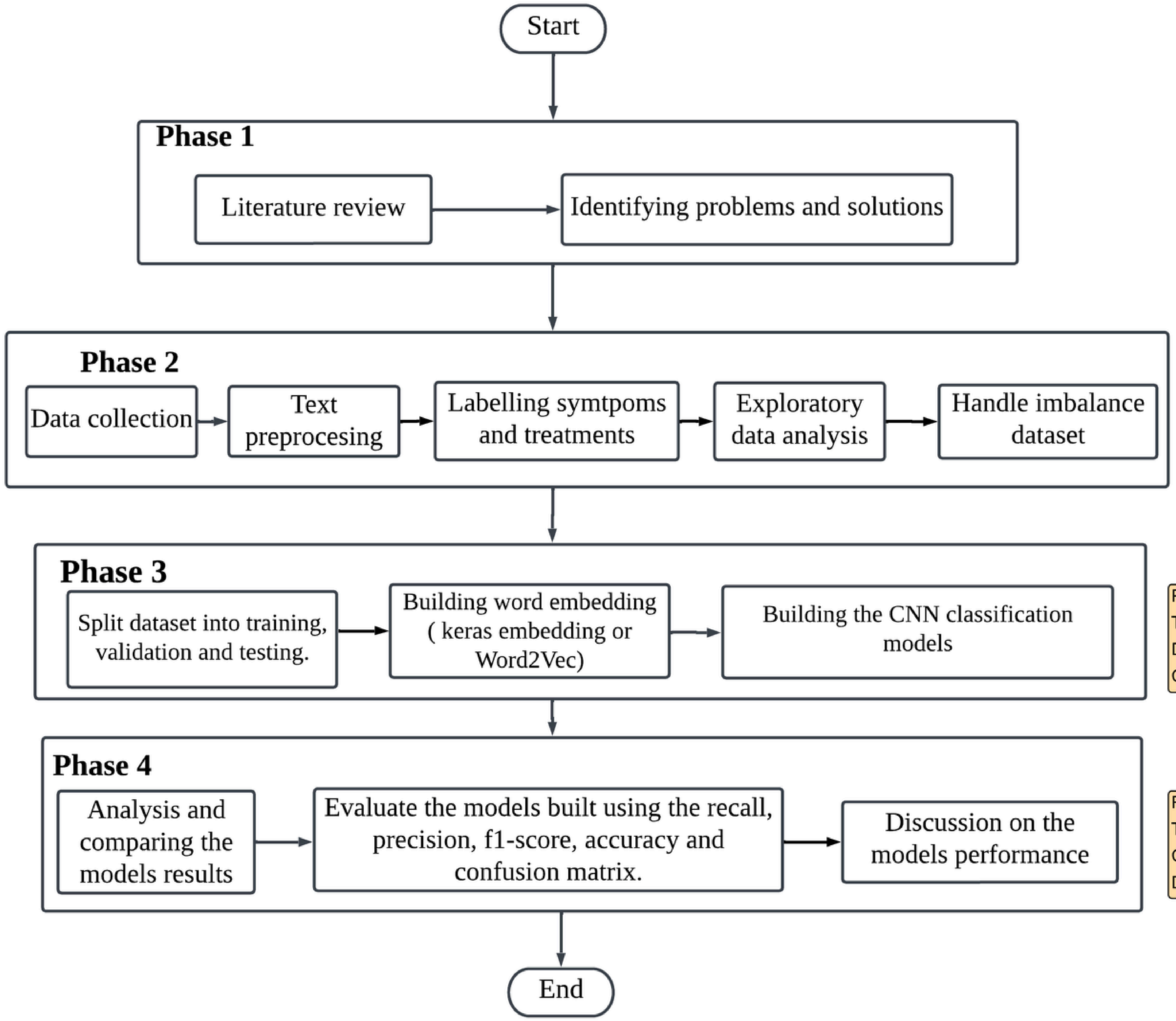
Why not combine the recent research ideas?

Conclusion: Research on text classification for MDD symptoms and treatments using CNN algorithms is able to be conducted.

Back to table of contents

CHAPTER 3 RESEARCH METHODOLOGY

FLOW OF RESEARCH FRAMEWORK



Research Objective 1:
To identify the related features for Major Depression Disorder (MDD) symptoms and treatments in the medical journals.

Research Objective 2:
To perform the text classification method for collection of Major Depression Disorder (MDD) symptoms and treatments using Convolutional Neural Networks (CNN) algorithms.

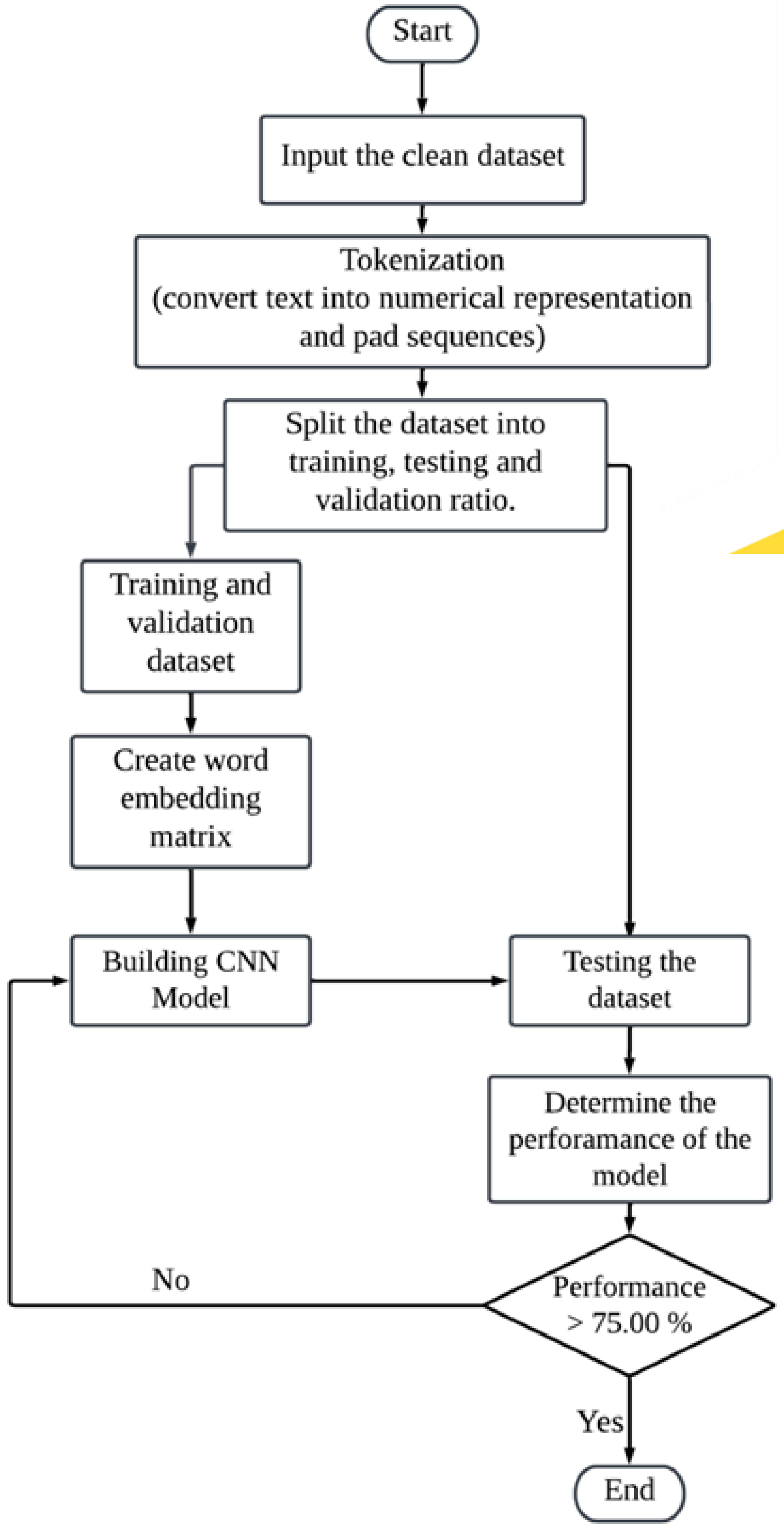
Research Objective 3:
To evaluate the performance of the machine learning methodologies, Convolutional Neural Networks (CNN) in order to identify the Major Depression Disorder (MDD) symptoms and treatments.

Back to table of contents

FLOW OF CNN MODEL BUILDING

3 CNN Model will be built:

- Benchmark CNN Model
- Proposed CNN Model
- Proposed CNN Model + Word2Vec



There are 3 sets:

- Set 1:
  - 80% Training
  - 10% Testing
  - 10% Validation
- Set 2:
  - 70% Training
  - 15% Testing
  - 15% Validation
- Set 3:
  - 60% Training
  - 20% Testing
  - 20% Validation

Back to table of contents

CHAPTER 4
RESEARCH DESIGN & IMPLEMENTATION

Back to table of contents

DATASET PREPARATION

1. Extract medical journals from NCBI website
2. Split abstracts into sentences
3. Text Preprocessing
4. Labelling Symptoms and Treatments
5. Exploratory Data Analysis
6. Handle Imbalanced Dataset

Back to table of contents

STAGE 3: TEXT PREPROCESSING

Text Preprocessing

- Removal of stop words
- Removal of digits
- Removal of punctuation
- Removal of empty row
- Convert text to lowercase
- Tokenization
- Lemmatization

```
sentences_df.head(10)
```

	sentences
0	[StringElement('Postpartum depression (PPD) is...
1	China is planning to launch PPD screening in c...
2	', attributes={'Label': 'BACKGROUND', 'NlmCate...
3	Video structured diagnostic interviews were pe...
4	Optimal screening was determined based on the ...
5	', attributes={'Label': 'METHODS', 'NlmCategor...
6	Among those screened, the video structured dia...
7	The optimal screening approach involved combin...
8	', attributes={'Label': 'RESULTS', 'NlmCategor...
9	', attributes={'Label': 'LIMITATIONS', 'NlmCat...

```
sentences_df.head(10)
```

	sentences
0	stringelementpostpartum depression ppd importa...
1	china planning launch ppd screening community ...
2	attributeslabel background nlmcategory backgro...
3	video structured diagnostic interview performe...
4	optimal screening determined based acceptabili...
5	attributeslabel method nlmcategory method stri...
6	among screened video structured diagnostic int...
7	optimal screening approach involved combining ...
8	attributeslabel result nlmcategory result stri...
9	attributeslabel limitation nlmcategory conclus...

Back to table of contents

SUMMARY OF KEYWORDS

Authorised Official Websites	Symptoms	Treatments
World Health Organization (WHO)	• poor concentration • feelings of excessive guilt or low self-worth • hopelessness about the future • thoughts about dying or suicide • disrupted sleep • changes in appetite or weight • feeling very tired or low in energy.	• antidepressants • behavioural activation • cognitive behavioural therapy • interpersonal psychotherapy • problem-solving therapy • Psychotherapy • fluoxetin
VeryWell Health	• Persistent sadness, anxious, or “empty” mood • Feelings of hopelessness, or pessimism • Irritability • Feelings of guilt, worthlessness, or helplessness • Loss of interest or pleasure in hobbies and activities • Decreased energy or fatigue • Moving or talking more slowly • Feeling restless or having trouble sitting still • Difficulty concentrating, remembering, or making decisions • Difficulty sleeping, early-morning awakening, or oversleeping • Appetite and/or weight changes • Aches or pains, headaches, cramps, or digestive problems without a clear physical cause and/or that do not ease even with treatment • Thoughts of death or suicide, or suicide attempts	• Persistent sadness, anxious, or “empty” mood • Feelings of hopelessness, or pessimism • Irritability • Feelings of guilt, worthlessness, or helplessness • Loss of interest or pleasure in hobbies and activities • Serotonin-norepinephrine reuptake inhibitors (SNRIs) • Transcranial magnetic stimulation (TMS) • Tricyclic antidepressants

Back to table of contents

SUMMARY OF KEYWORDS

Authorised Official Websites	Symptoms	Treatments
Mayo Clinic	• Loss of interest or pleasure • Insomnia • Headaches • Feelings of guilt, worthlessness	• Atypical antidepressants • Electroconvulsive therapy (ECT) • fluoxetine • Monoamine oxidase inhibitors (MAOIs) • neuromodulation • Selective serotonin reuptake inhibitors (SSRIs) • Serotonin-norepinephrine reuptake inhibitors (SNRIs) • Transcranial magnetic stimulation (TMS) • Tricyclic Antidepressants
DSM-5 diagnostic manual	• depressed mood • loss of interest or pleasure; • weight loss or gain; • insomnia or hypersomnia • psychomotor agitation or retardation • fatigue • feelings of worthlessness or excessive guilt • decreased concentration • thoughts of death or suicide.	

Back to table of contents

STAGE 4: LABELLING OF SYMPTOMS AND TREATMENTS

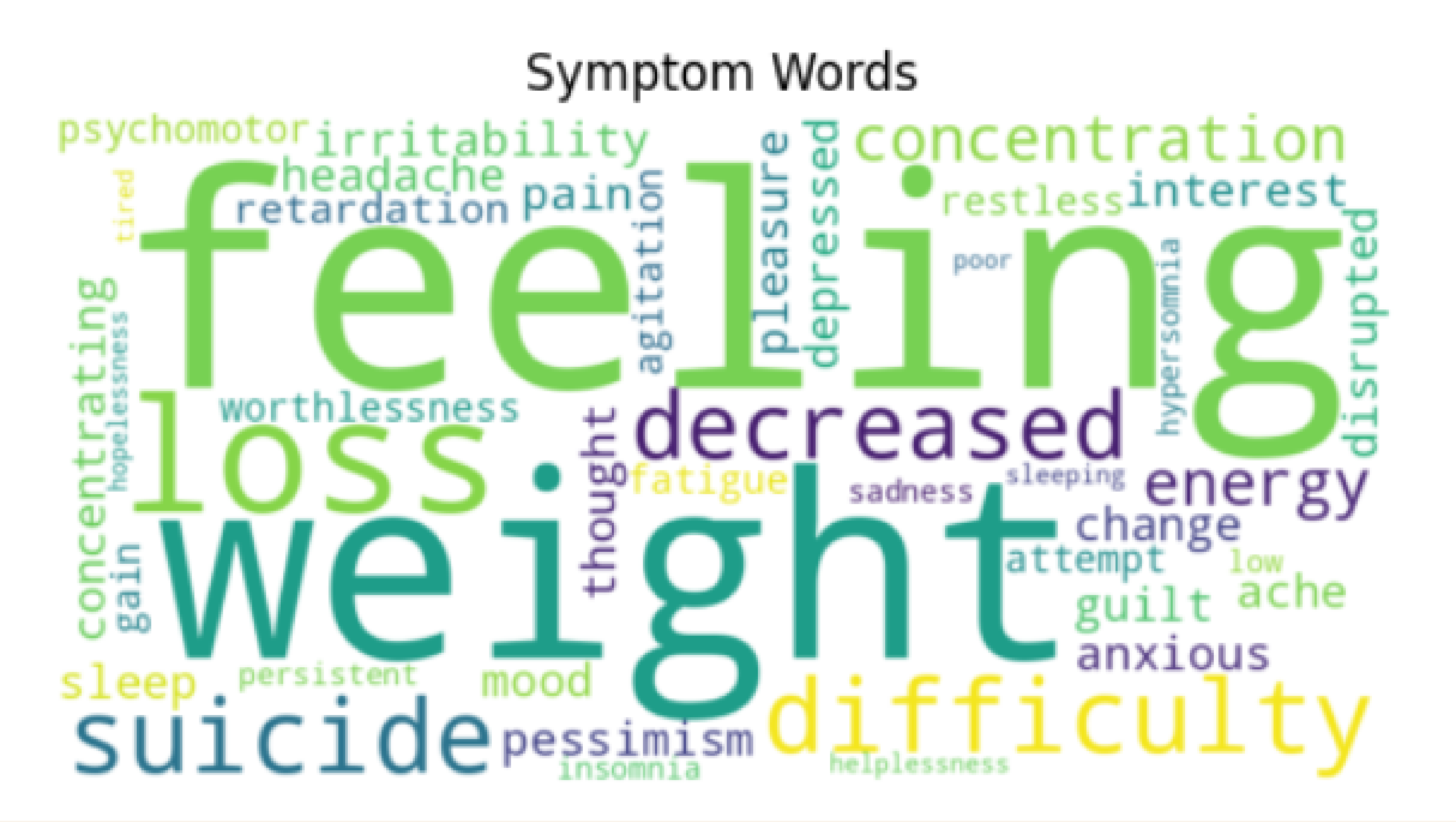
Labelling Process

- There are 48 keywords defined for symptoms and 19 keywords for treatments which retrieved from the four authorised websites.
- Keywords will undergo text preprocessing.
- The sentences are labelled based on matching the keywords defined.
- Matching the keywords gram by gram in the sentences by using the ngram in NLTK and loops.
- There are 4 labels:
  - Label ‘0’: represent ‘none’, no symptoms and treatments found in the sentences.
  - Label ‘1’: represent there is any symptoms word found in the sentences.
  - Label ‘2’: represent there is any treatments word found in the sentences.
  - Label ‘3’: represent ‘both’ symptoms and treatments word found in the sentences.

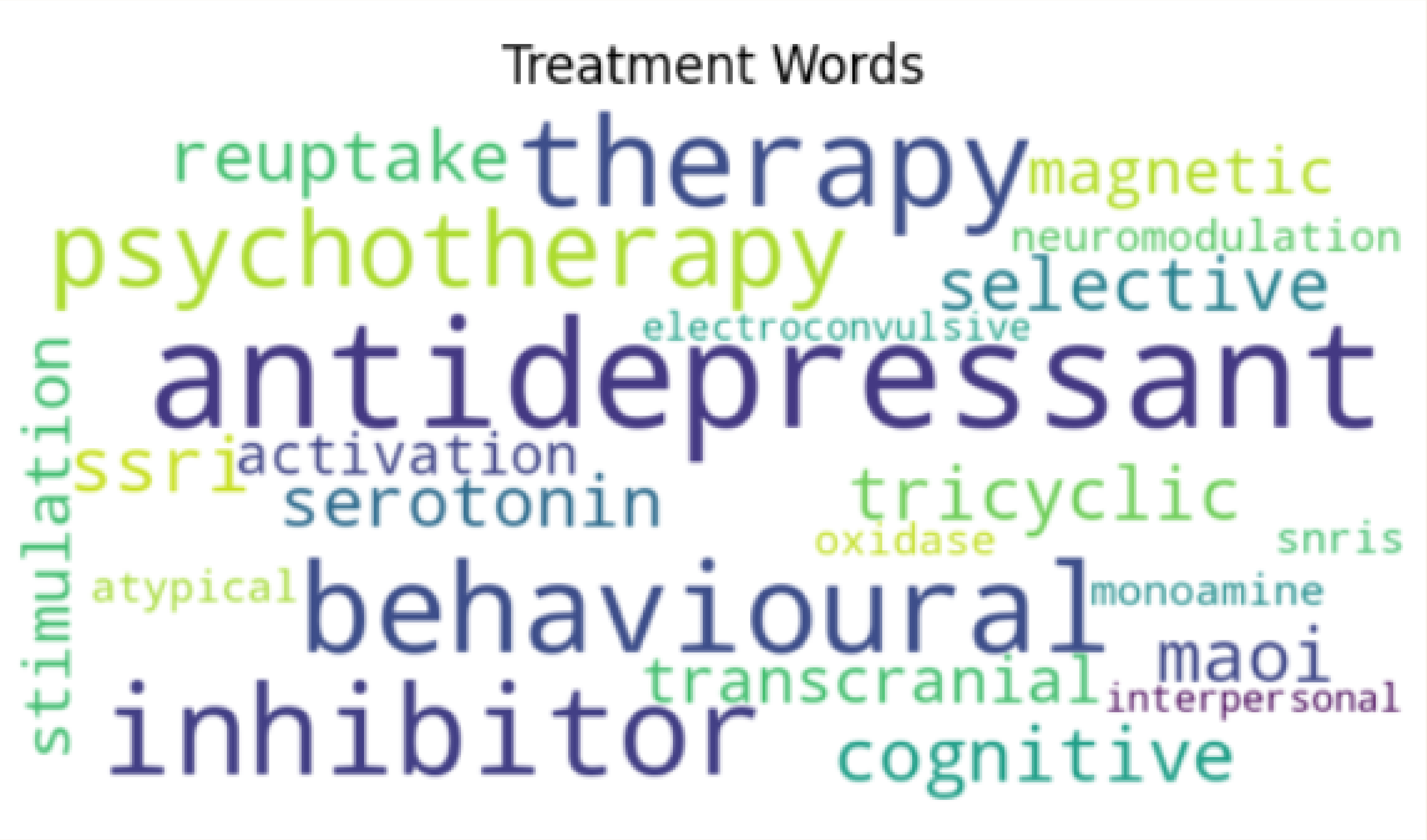
Back to table of contents

STAGE 5: EXPLORATORY DATA ANALYSIS

Word Cloud for Symptoms

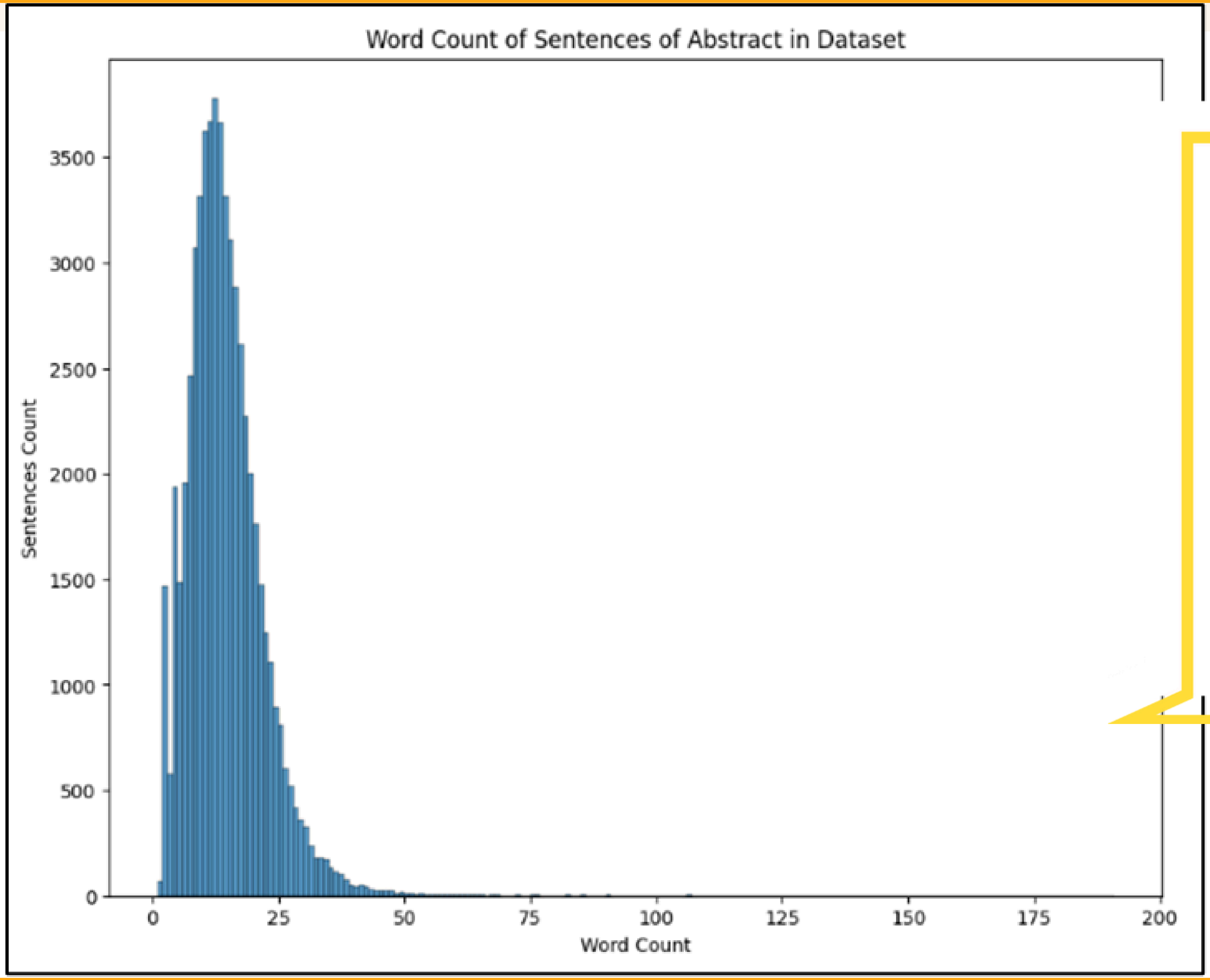


Word Cloud for treatments



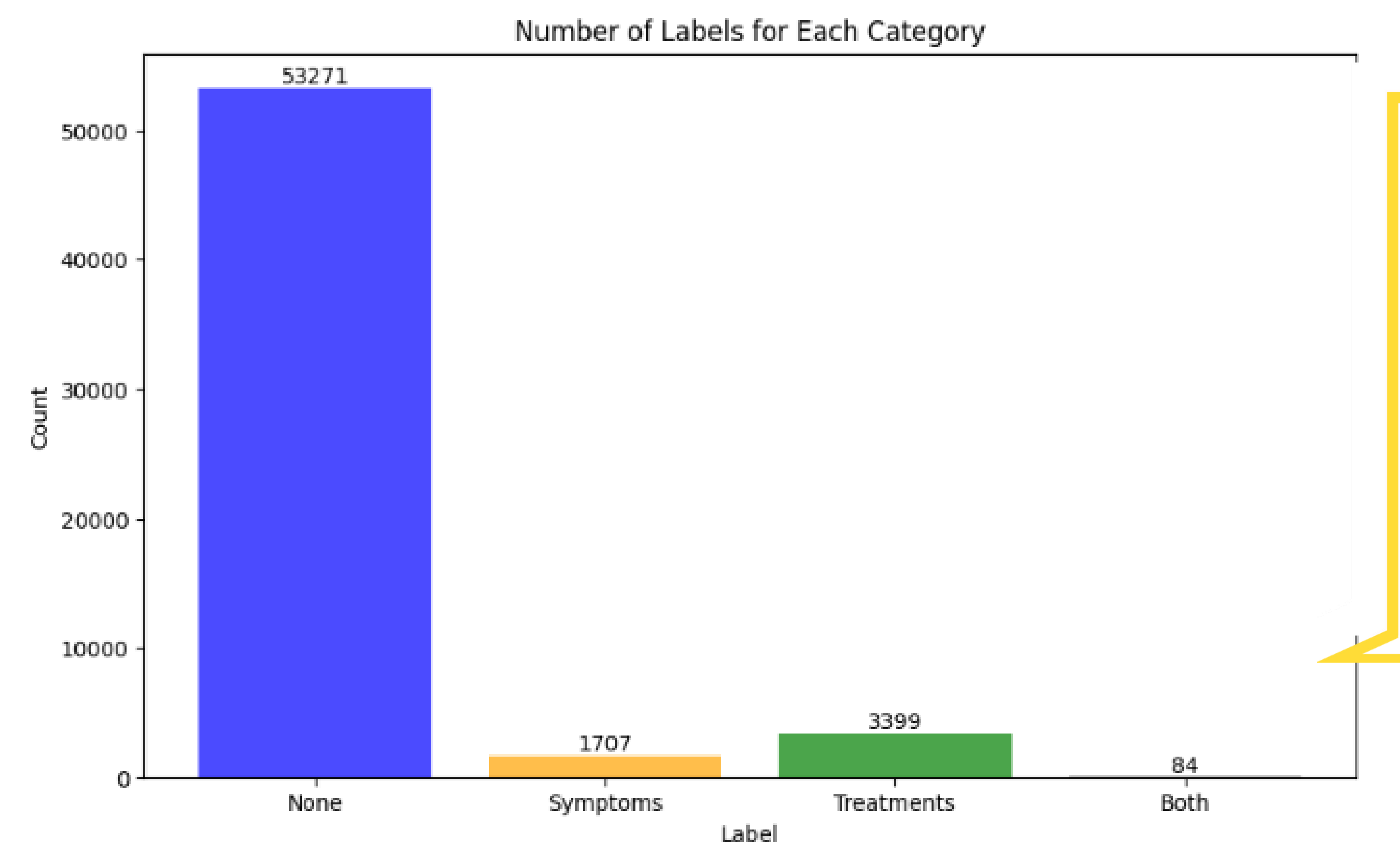
STAGE 5: EXPLORATORY DATA ANALYSIS

Word Count of Sentences of Abstract in Dataset



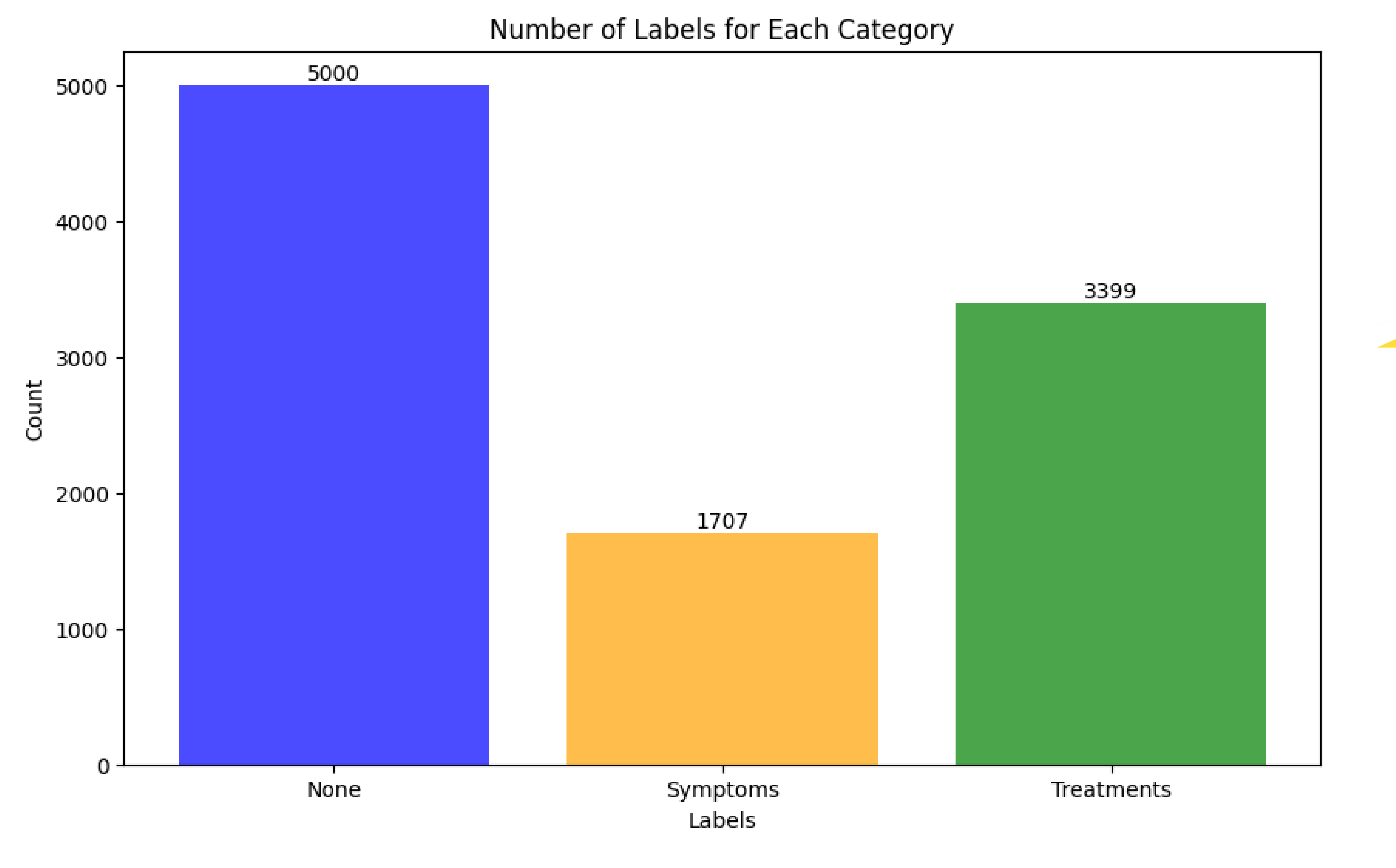
To identify the maximum length of words in the datasets per sentences

STAGE 5: EXPLORATORY DATA ANALYSIS



- the 'None' category will undergo undersampling (resampling technique) to be reduced to 5000.
- For the 'both' category, it will be removed since the amount of values is too small and it does not have significant impact for the model to achieve the research's objectives.

STAGE 6:HANDLE IMBALANCED DATASET



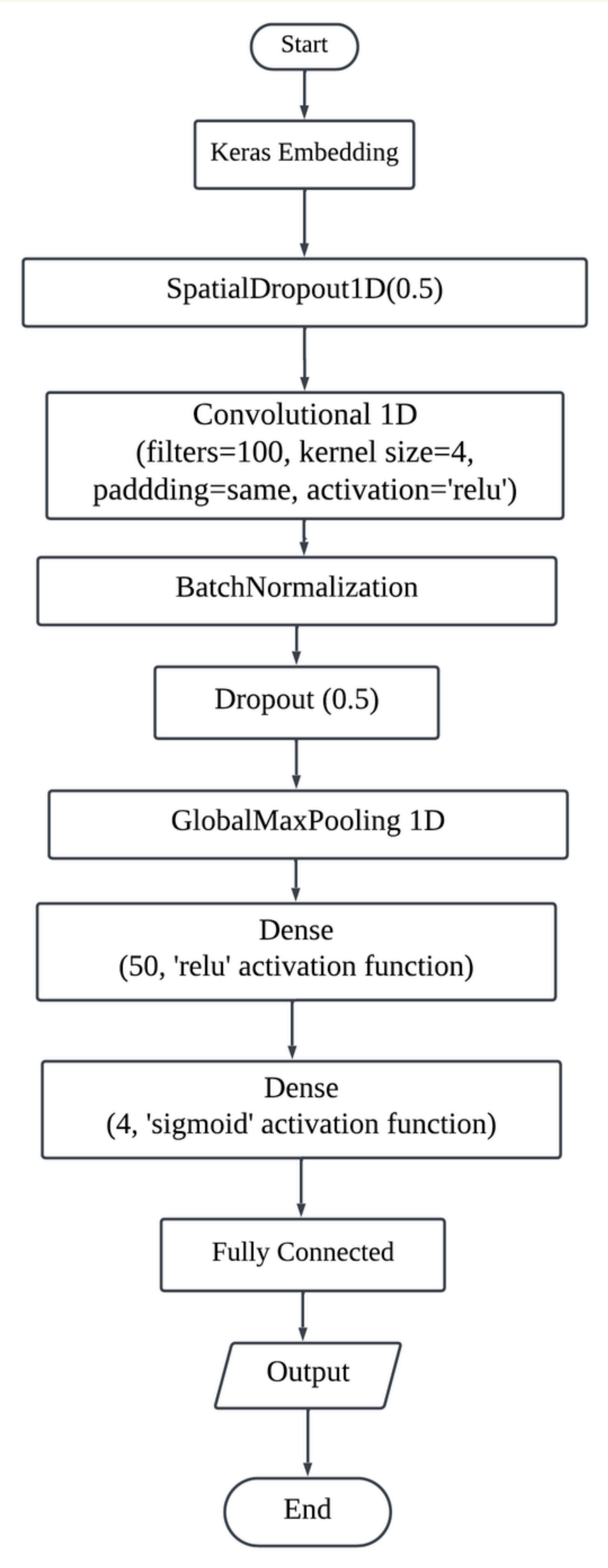
Final Dataset

CNN MODEL BUILDING

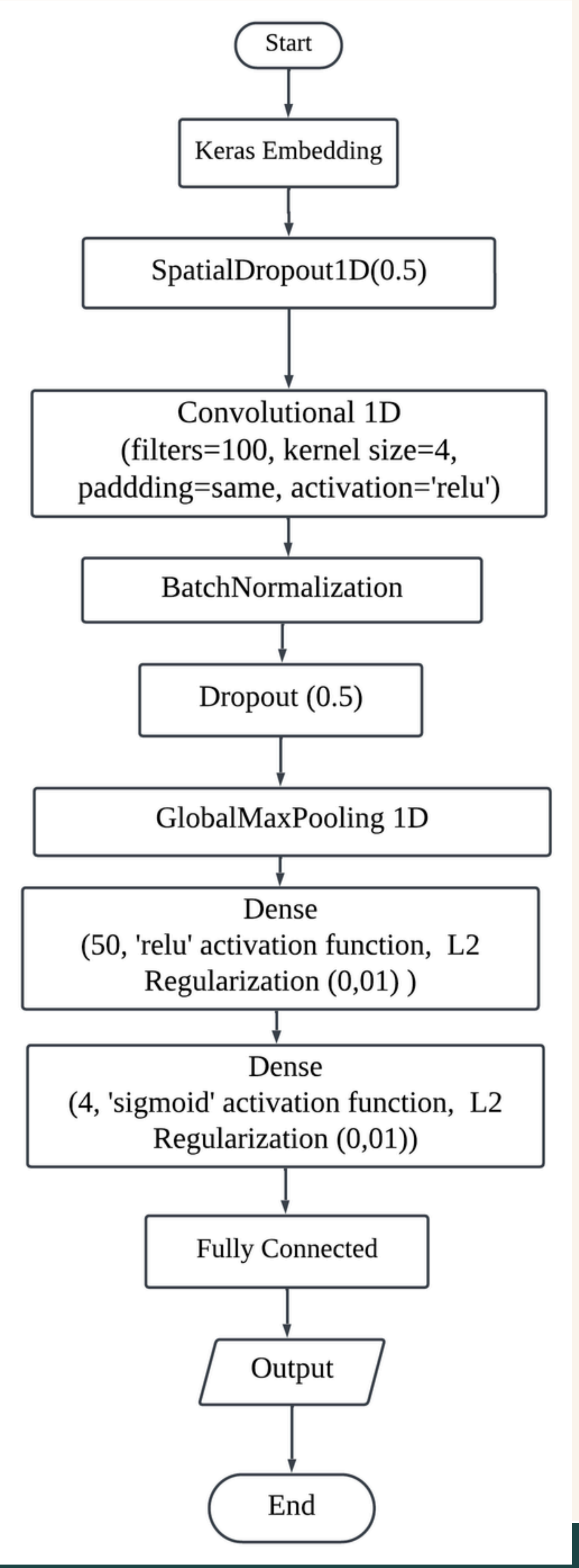
	Benchmark CNN Model	Proposed CNN Model	Proposed CNN Model+Word2Vec
Embedding Layer	• Keras Embedding Layer	• Keras Embedding Layer	• Word2vec Embedding Layer
Learning Rate	• 0.01	• 0.001	• 0.001
Regularization Techniques Used	• Spatial Dropout Layer • Dropout layer	• Spatial Dropout Layer • Dropout layer • L2 Regularization	• Spatial Dropout Layer • Dropout layer • L2 Regularization
Activation Function	• ReLU (Rectified Linear Unit) • Sigmoid		
Loss Function	• Sparse Categorical Cross Entropy		

Back to table of contents

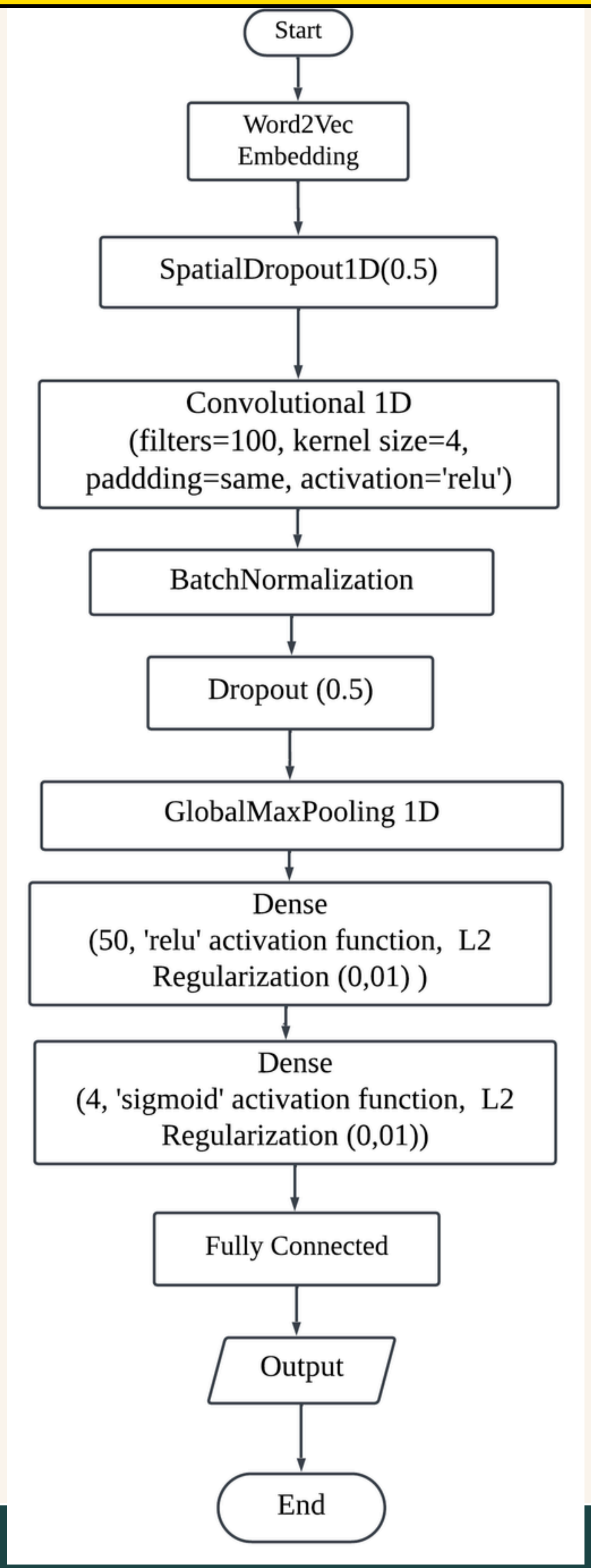
Benchmark CNN Model



Proposed CNN Model



Proposed CNN Model + Word2vec



Back to table of contents

CHAPTER 5
RESULTS ANALYSIS AND DISCUSSION

Back to table of contents

DATASET SPLITTNG RATIO RESULTS

Models	Accuracy, None (Precision, Recall, F1-Score), Symptoms (Precision, Recall, F1-Score) & Treatments (Precision, Recall, F1-Score)		
	Set 1	Set 2	Set 3
Benchmark CNN Model	• 0.91 • (0.87, 0.98, 0.92) • (0.97, 0.56, 0.71) • (0.97, 0.96, 0.97)	• 0.93 • (0.88, 0.99, 0.93) • (0.98, 0.65, 0.78) • (1.00, 0.95, 0.97)	• 0.88 • (0.87, 0.91, 0.89) • (0.74, 0.65, 0.69) • (0.97, 0.96, 0.96)
Proposed CNN Model	• 0.92 • (0.88, 0.97, 0.92) • (0.91, 0.63, 0.74) • (0.98, 0.96, 0.97)	• 0.92 • (0.89, 0.96, 0.92) • (0.93, 0.67, 0.78) • (0.96, 0.97, 0.97)	• 0.91 • (0.88, 0.95, 0.92) • (0.89, 0.68, 0.77) • (0.97, 0.96, 0.96)
Proposed CNN Model + Word2Vec	• 0.92 • (0.88, 0.98, 0.92) • (0.93, 0.61, 0.74) • (0.98, 0.96, 0.97)	• 0.92 • (0.88, 0.97, 0.93) • (0.94, 0.66, 0.77) • (0.98, 0.97, 0.97)	• 0.92 • (0.88, 0.97, 0.92) • (0.95, 0.66, 0.78) • (0.97, 0.96, 0.96)

Back to table of contents

COMPARISON BETWEEN DATASET SPLITTNG RATIO RESULTS

Models	Comparison between Set 1, Set 2 and Set 3
Benchmark CNN Model	• Set 2 has the highest accuracy (0.93), followed by set 1 (0.91) and set 3 (0.88). • Set 2 has the highest F1-score (0.78) for symptoms label. • Set 1 and Set 2 has the highest F1-score (0.97) for treatments label.
Proposed CNN Model	• Set 1 and Set 2 has the highest accuracy (0.92). • Set 2 has the highest precision (0.93) & F1-score (0.78) for symptoms label. • Set 1 and Set 2 has the highest F1-score (0.97) for treatments label.
Proposed CNN Model + Word2Vec	• All sets have accuracy of 0.92. • Set 3 has the highest precision (0.95) & F1-score (0.78) for symptoms label. • Set 1 and Set 2 has the highest F1-score (0.97) for treatments label. • Set 2 has the highest recall (0.97) for treatments label.

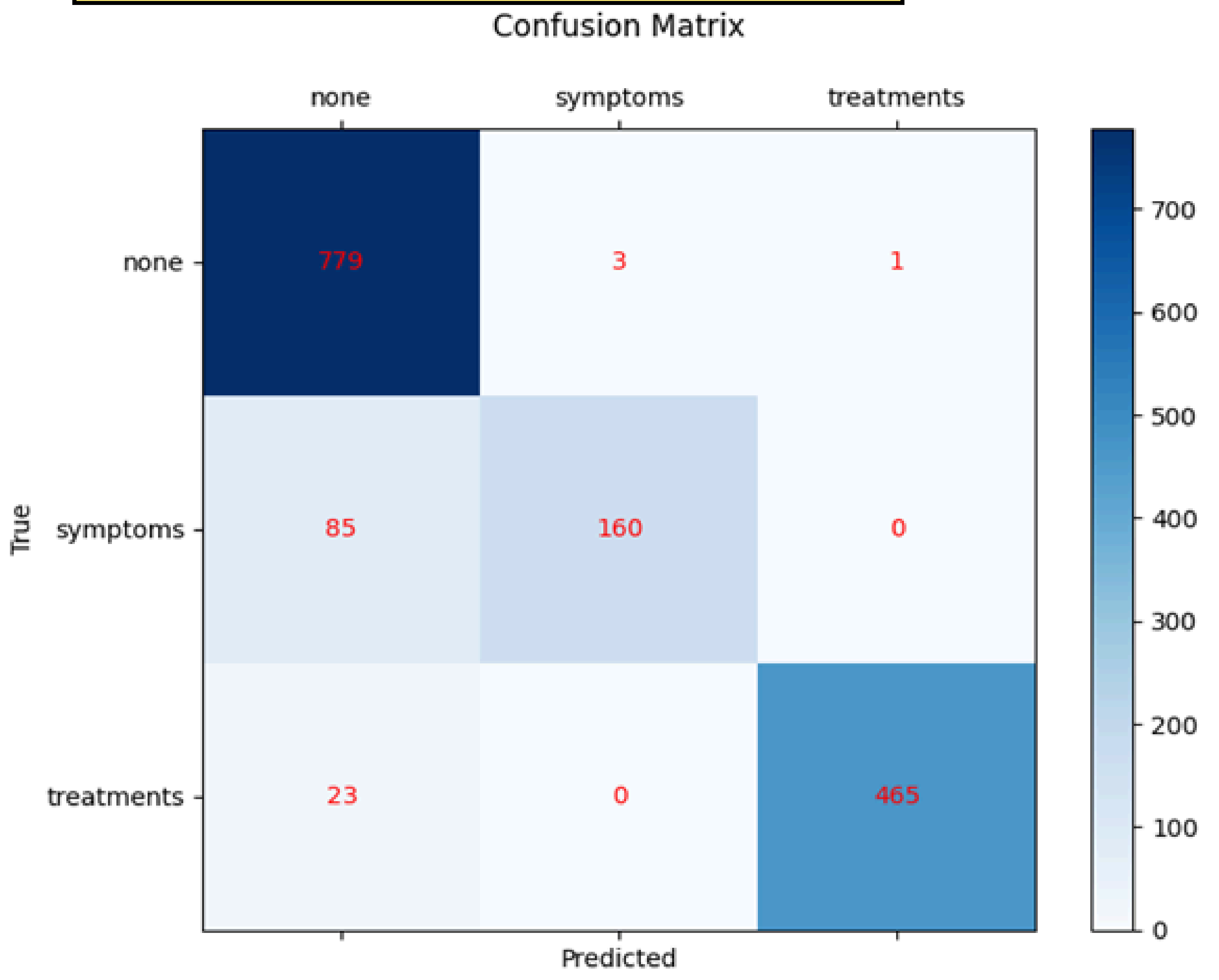
There are 3 sets:
- Set 1:
  - 80% Training
  - 10% Testing
  - 10% Validation
- Set 2:
  - 70% Training
  - 15% Testing
  - 15% Validation
- Set 3:
  - 60% Training
  - 20% Testing
  - 20% Validation

Back to table of contents

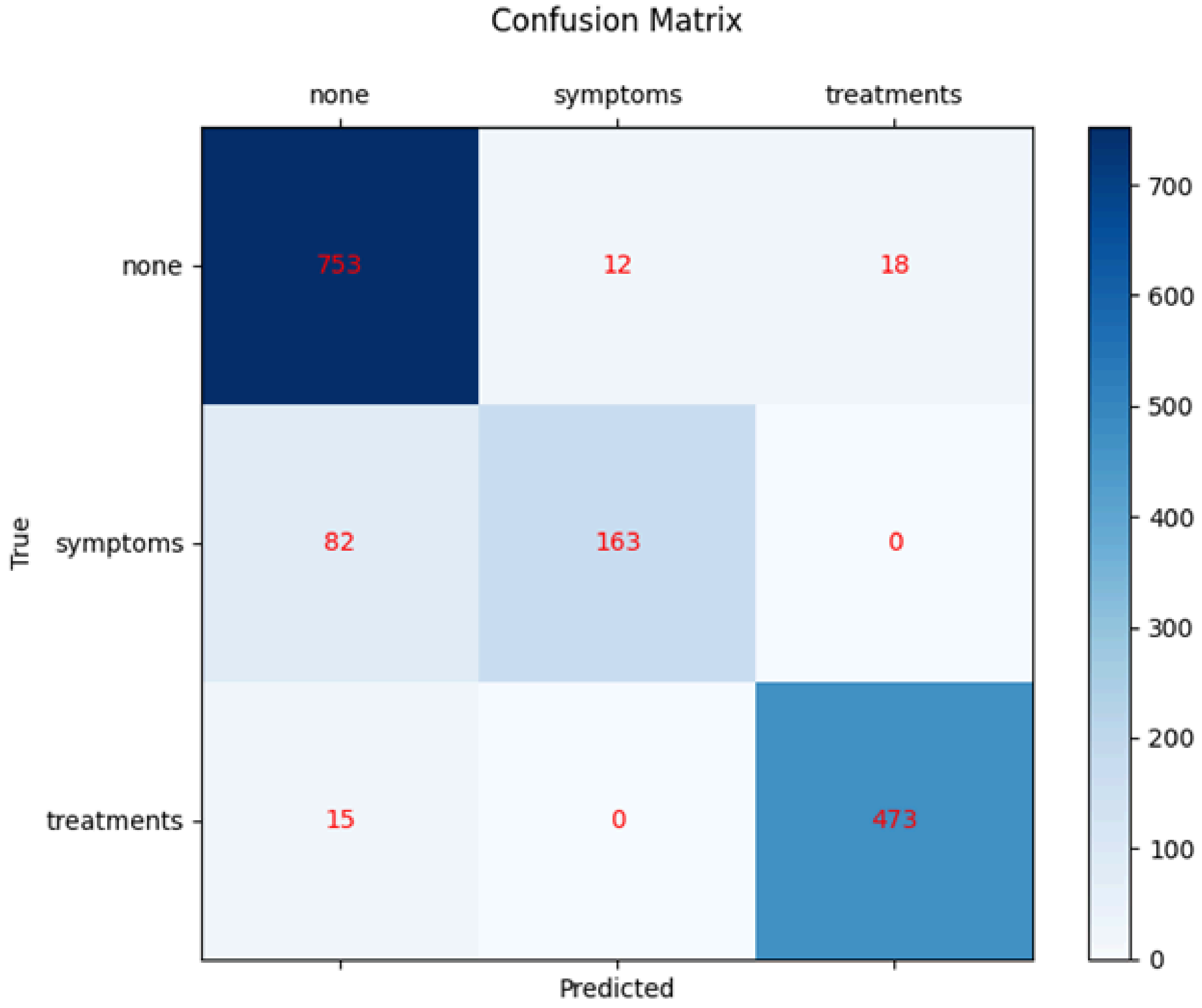
CONFUSION MATRIX RESULTS

Results

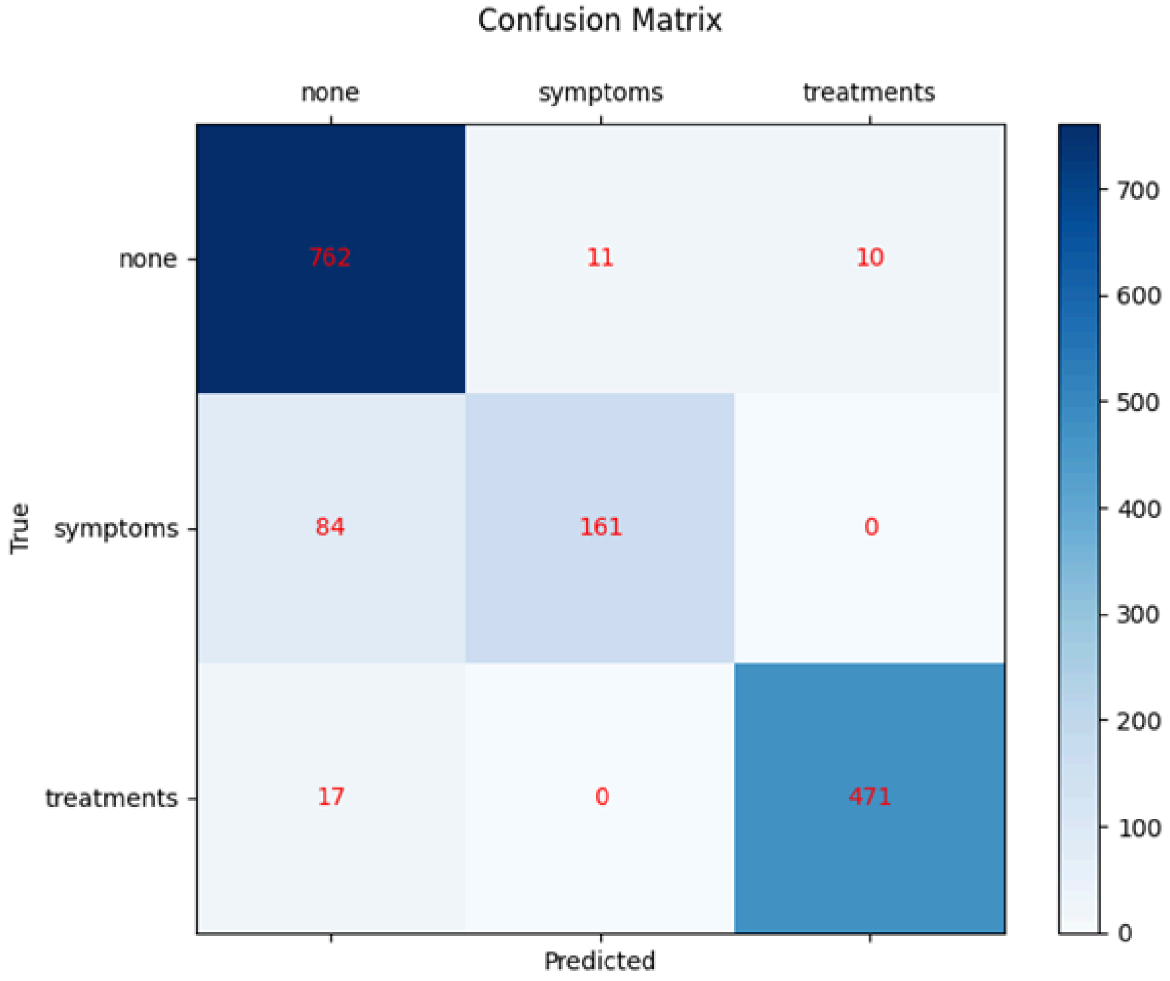
Benchmark CNN Model



Proposed CNN Model



Proposed CNN Model + Word2vec



Back to table of contents

RESULTS FOR 3 MODELS

Models	Accuracy	None (Precision, Recall, F1-Score)	Symptoms (Precision, Recall, F1-Score)	Treatments (Precision, Recall, F1-Score)
Benchmark CNN Model	0.93	• 0.88 • 0.99 • 0.93	• 0.98 • 0.65 • 0.78	• 1.00 • 0.95 • 0.97
Proposed CNN Model	0.92	• 0.89 • 0.96 • 0.92	• 0.93 • 0.67 • 0.78	• 0.96 • 0.97 • 0.97
Proposed CNN Model + Word2Vec	0.92	• 0.88 • 0.97 • 0.93	• 0.94 • 0.66 • 0.77	• 0.98 • 0.97 • 0.97

Back to table of contents

COMPARISON BETWEEN 3 MODELS

Models	Benchmark CNN Model	Proposed CNN Model	Proposed CNN+ Word2vec
Accuracy	Has the highest accuracy (0.93).	Accuracy drop by 1% (0.92)	Same with proposed model, does not have significant impact.
None (Precision, Recall & F1-Score)	Has the highest recall (0.99).	Precision increased by 1% (0.89), but recall (0.96) and F1-score (0.92) decreased.	Precision same with benchmark model (0.88), recall (0.97) and F1-score (0.93) slightly increased.
Symptoms (Precision, Recall & F1-Score)	Has the highest precision (0.98) and F1-Score (0.78).	Precision decreased by 5% (0.93), but recall increases by 2% (0.67), have same F1-score with benchmark model.	Precision increased by 1% (0.94), but recall (0.66) and F1-score (0.77) slightly decreased.
Treatments (Precision, Recall & F1-Score)	Overfitting occurs due to 100% in precision.	Precision value decreased to 0.96 after using the L2 regularization and reduce the learning rate from 0.01 to 0.001.	Precision (0.98) values increased by 2%.

Back to table of contents

CHAPTER 6
CONCLUSION & RECOMMENDATIONS

CONCLUSION

- All three models perform well when using the set 2 dataset split ratio.
- Proposed CNN Model is outperformed than other models when using set 2 dataset split ratio even though Benchmark CNN Model has the highest accuracy, 93.00% but with overfitting occurs.
- Proposed CNN Model is the best model that classifies the MDD symptoms and treatments using the medical journal articles.

ACHIEVEMENTS

- MDD symptoms and treatments in the datasets were successfully labelled based on the predefined keyword.
- Three CNN models are built to classify the MDD symptoms and treatments in the datasets.
- Performance of three CNN models had evaluate and each models have accuracy more than 90.00%.

Back to table of contents

FUTURE WORKS

- Identify solutions to increase the corpus of keywords of MDD symptoms and treatments to provide a more precise labelling process.
- Implement other methods to handle imbalanced datasets rather than undersampling techniques.
- Experiment with different embedding layers other than Keras and Word2vec to enhance the performance of the models.
- Perform hyperparameter tuning to enhance the model's performance.

Back to table of contents


THE END