ABSTRACT

In light of the discovery of the SARS-CoV-2 virus, the scientific community has put forth an extraordinary effort that has resulted in the advancements of over 300 vaccine concepts. Over 40 are being evaluated in clinical studies; some are in Phase three trials, while some have completed Phase three with positive findings. Several of them novel vaccines have previously been approved for use in emergency situations. Furthermore, whether and to what degree the capacity of the vaccinations under consideration, as well as unrelated vaccines, can improve immunological fitness.

Because of the short making time and the newness of the technology used, these vaccines are released with a number of unresolved problems that will only be cleared with time. Technical issues relating to the production of tons of doses, as well as ethical issues relating to the availability of these vaccinations in even the lower economy countries, are looming concerns. In order to achieve equal global availability, protection of various domains, and immunisation against virus variations, we believe that more than one vaccination will be required in the long term. The goal of this work is to identify the potential for repurposing medications by utilising the covid-19 articles to make it easier for researchers. To capture the sentiments of Top Drug Names who have given Covid-19 therapies a positive review.

TABLE OF CONTENTS

CHAPTER 1 **Page No.**

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

LIST OF TABLES

Table No. **Title of the Figure** **Page No.**

LIST OF FIGURES

Fig No. **Title of the Figure** **Page No.**

CHAPTER 1

1.1 INTRODUCTION

Drug development is a time consuming, costly , and high risk process, drug repurposing has become popular (also known as drug repositioning or drug re-profiling). Furthermore, it is unknown if and to what extent the vaccines under evaluation, as well as associated immunizations, can improve immunological fitness by training innate immunity against SARS-CoV-2 and pathogenesis. Furthermore, whether and to what degree the capability of vaccines under assessment and unrelated vaccinations can boost immunological fitness by training innate immunity to SARS-CoV-2 and pathogenesis remains unknown.

The COVID-19 epidemic has made rapid advancements in research on the subject a global necessity. During this time, the amount of data created far outstripped human analysts' ability to keep up with it. These vaccines will be launched with various unresolved difficulties due to the short development time and novelty of the technology used.

Technical issues relating to the production of billions of doses, as well as ethical issues relating to the availability of these vaccinations in even the poorest countries, are looming concerns.

In order to achieve equal global availability, protection of various subjects, and immunisation against virus variations, we believe that more than one vaccination will be required in the long term. Aspirin is a well-known example, which was first used for pain eradication and is now used to prevent cardiovascular disease and other illness like cancer. In the fight against COVID-19, drug repositioning is a hot topic in biomedical research, and it's happening at the same time as vaccine research and development.

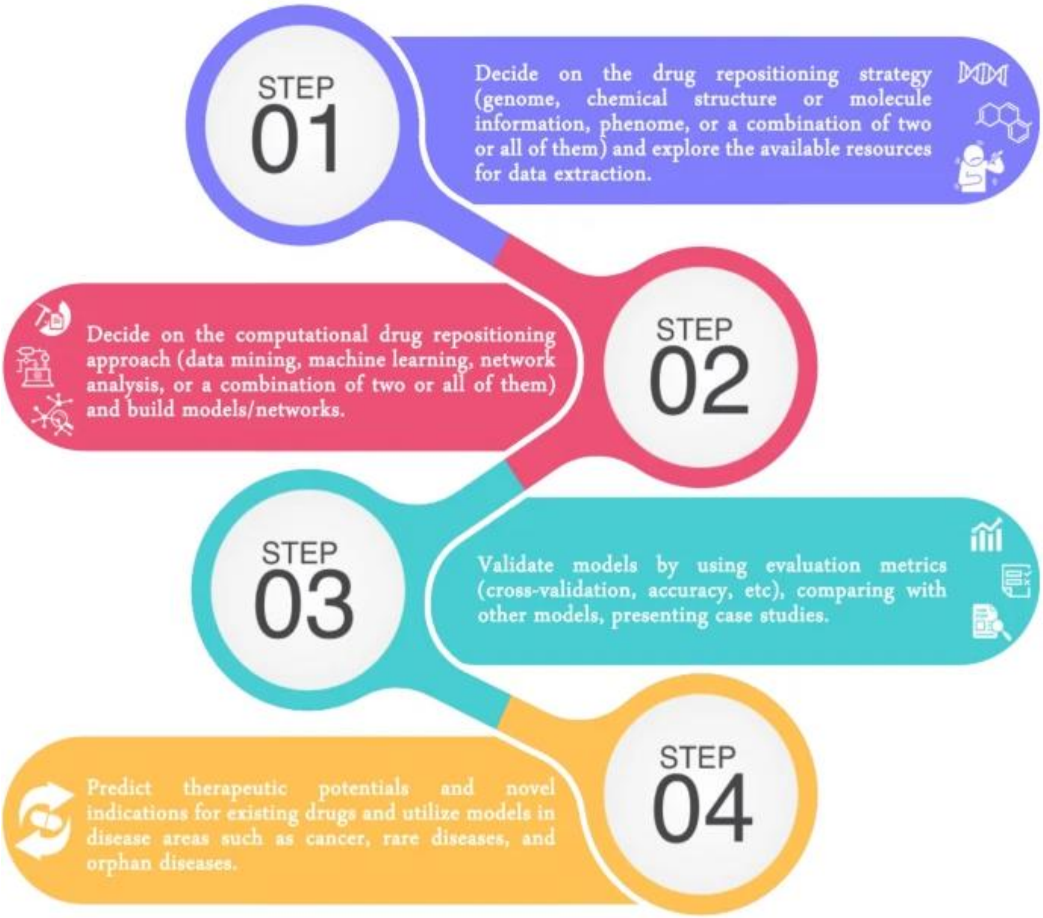
As one of the most active fields in pharmacology in the previous decade, drug repurposing is a major industry with a bright future. Drugs are continuously being studied to see whether they can be used for new reasons that they weren't designed for. When compared to traditional procedures, it may enable for more systematic and significantly less cost ways in the identification of new medicines for diseases.

Multidisciplinary academics and scientists have made various attempts to computationally analyse the potential of repositioning pharmaceuticals to uncover alternate medicinal indications, with varying degrees of efficiency and success.

Traditionally, drug repositioning studies have centred on identifying drug reaction and mode of action relatedness, identifying novel drug indications by screening the current pharmacopoeia against new targets, investigating common characteristics between drug compounds such as chemical structures and side effects, and discovering drug-disease relationships.

Large-scale biomedical and electronic health-related data, such as microarray gene expression signatures, pharmaceutical databases, and online health communities, has accelerated the development of computational drug repositioning approaches, which typically include text mining, automated learning.

The study of the link between diverse biomedical components is an important part of modern drug repositioning research. Drugs, genes, and adverse drug reactions are examples of biological entities.



STEP 01

Decide on the drug repositioning strategy (genome, chemical structure or molecule information, phenome, or a combination of two or all of them) and explore the available resources for data extraction.

STEP 02

Decide on the computational drug repositioning approach (data mining, machine learning, network analysis, or a combination of two or all of them) and build models/networks.

STEP 03

Validate models by using evaluation metrics (cross-validation, accuracy, etc), comparing with other models, presenting case studies.

STEP 04

Predict therapeutic potentials and novel indications for existing drugs and utilize models in disease areas such as cancer, rare diseases, and orphan diseases.

Fig. 1.1 Drug Repurposing Flow Diagram

Drug repurposing has been shown to be a promising strategy to improve drug development with these repurposed medications. To uncover the possibility for repurposing medications, researchers used the covid-19 articles to make it easier for them.

Sentiment analysis is required. Sentiment analysis features can aid in the detection of adverse medication responses in text. To capture the sentiments of Top Drug Names using various NLP sentiment analysis techniques who have given Covid-19 therapies a positive review. Within the articles about covid-19 treatment, look for the most relevant drug names and display it using a word or cloud chart.

1.2 DATASET DETAILS

In the quest to find effective COVID19 treatments and management measures, CORD-19 wants to bring together the machine learning community with biomedical domain specialists and policymakers. Papers in CORD-19 come from PubMed Central (PMC), PubMed, the World Health Organization's Covid-19 Database4, and the bioRxiv, medRxiv, and arXiv preprint services as shown in Fig. 1.2.

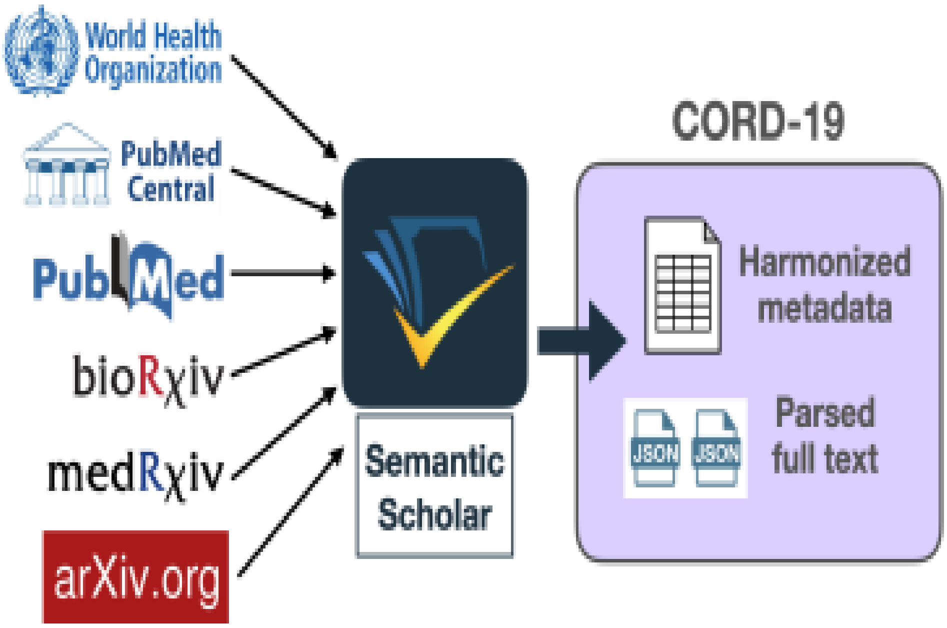


Fig. 1.2 CORD-19 Dataset

1.3 PROBLEM STATEMENT

Repositioning or Repurposing of Existing Drugs which was intended to serve other medical problems for Treatment of Covid-19 using various Articles published by utilizing various NLP methods and Deep Learning Algorithms.

1.4 OBJECTIVE

- To find the potential of repurposing the drugs, using the covid-19 articles in order to make it easy for the researchers.
- Sentiment analysis is required, features of sentiment analysis can be helpful in detecting adverse drug reactions within the text.
- To capture the sentiments of Top Drug Names with positive notation for Covid-19 treatments.
- Find most relevant drug names within the articles related to covid-19 treatment.

CHAPTER 2

2.1 LITERATURE SURVEY

Several works reported focused on the literature information retrieval of covid-19 [1] “Literature-based discovery of new candidates for drug repurposing” published by Hsih-Te Yang and Jiun-Huang Ju, discusses to get illness gene direct correlations from the articles, we used a drug vector space model and a pattern-based relationship extraction approach. The chord diagram was used to integrate and summarise repurposed medication candidates for the treatment of different diseases, as well as additional signs, which was initially used to represent the cycle of drug repurposing. Saffron is an information extraction method that uses mast to recognise essential ability ideas in the space IDF [2], “Term Extraction Approach to Expert Finding on the COVID-19 Open Research Dataset” published by Cécile Robin et al. addresses discussed how to deal with the problem of restricted vocabularies and physically commented on distributions that fall behind while looking for late-arriving subjects and grounded scient metric techniques are notoriously difficult to analyse across logical regions and applications.

“Interactive Visualization and Simplified Pattern Discovery in the COVID-19 Open Research Dataset (CORD-19) [3]” published by Wuraola Fisayo Oyewusi focuses on Step-by-step directions for visualising a large dataset and how to think about examples, relationships based on frequency of disease and compound elements using Scispacy. Scispacy Named Entity Recognition Model (en ner bc5cdr md) based on the Bio Creative V Chemical-Disease Relations (BC5CDR) where Each of the information points [4] in the content perception, when clicked, presents the where each of the information point in the content perception, when clicked shows the content information in various settings from the CORD-19 dataset. There is additionally a hunt box where clients of this device can type in words identified with sicknesses and synthetic substances, they think might be huge to their examination. "A Systematic Framework for Drug Repurposing based on Literature Mining" authored by Gaocai Dong et al., explains the ABC model was used to develop an analogical reasoning technique and a framework for uncovering disease drug connections [5]. Two vectors are used to explain clinical illness symptoms and clinical medication effects. A logistic regression modelling is used to predict how good the drug emerges the disease related to the illness/drug vector that were created.

"DeepH-DTA: Deep Learning for Predicting Drug-Target Interactions: A Case Study of COVID-19 Drug Repurposing" authored by Mohamed Abdel Basset and Hossam Hawash, The model of heterogeneous graph attention (HGAT) is discussed in relation to learning topological information about complex molecules. (SMILING) To learn the Simplified Molecular Input Line Entry System representation [6] of input molecules, the bidirectional ConvLSTM architecture is used. A framework for estimating DT (drug target) binding affinity has been developed using target protein sequences.
Deep neural network models were studied from drug-review posts as a strategy for detecting unexpected drug consumption in social media as a two class classification challenge [7] . "Detecting Serendipitous Drug Usage in Social Media with Deep Neural Network Models" by Boshu Ru talks about Convolutional neural networks include the convolutional neural network, the long short-term memory and the convolutional memory network.

FNN uses link prediction, community understanding, graph depended learning, literature-based learning, and FNN to categorise pharmaceuticals into pharmaceutical therapeutic classes based on the drugs [8]. "An Analytical Review of Computational Drug Repurposing" by Seyedeh Shaghayegh Sadeghi and Mohammad Reza Keyvanpour addresses by Models based on publicly available biomedical literature data, these strategies depend on the analysis of openly available biomedical literature data to get indirect or intrinsic links between biological elements that appear to be unrelated [9].

“Semi-supervised graph cut algorithm for drug repositioning by integrating drug, disease and genomic associations” by Juan Liu addresses the drug ill bi graph with drug ill pairings as nodes, similarity between double pairs [10] represents the strength of the edge between two points using their similarity measurements. SSGC aims to forecast new drug disease therapy relationships by integrating three layers of data (treatment, gene, and base) [11]. “Term Extraction Approach to Expert Finding on the COVID-19 Open Research Dataset” published by Cécile Robin et al., addresses about How to take care of the issue of controlled vocabularies and physically commented on distributions that fall behind while looking for as of late arising subjects and grounded scient metric [12] approaches are famously hard to analyse across logical regions and utilizations Saffron information extraction system to distinguish important ability ideas in the space IDF and furthermore Implements master finding through democratic over naturally removed terms from text.

How can public specialists investigate pieces not unequivocally associated information worried to their individual obligations utilizing Semantic Textual Similarity (STS). SciBERT, BioBERT Infer Sent GloVe-840B [13], Infer Sent Glove SMT (for examination) Named-Entity Recognition and Disambiguation (NERD), gave by the Onto Gene's Biomedical Entity Recogniser (OGER), a cutting-edge biomedical annotator which thusly relies upon the Bio Term Hub (BTH), In this paper they had the option to do the Identification of focal things of information inside an assortment of distributions. SciBERT [14] was chosen over BioBERT on the grounds that COVID19 corpora incorporate writing from different regions other than natural sciences.

What are the elements realized through point demonstrating to choose a bunch of examination articles generally applicable to the jobs needing to be done utilizing the accompanying strategies like Non-negative lattice factorization (NMF) [15], TF-IDF, Key word coordinating. named element acknowledgment utilizing Scispacy en center sci sm language model. Discoveries of this paper incorporate Topic Modeling-based methodology gives incredible potential in serving in unaided, mechanized record recovery for a bunch of very much characterized errands [16]. Continual BERT, a novel BERT engineering based on existing methods to take in and remove outlines from a consistent stream of new undertakings while holding recently learned data. Bert Sum Elastic Weight Consolidation (EWC)- EWC ascertains the Bayesian [17] back circulation of boundaries utilizing Laplace estimate to align slope drop towards the covering learning district of both the past and new assignments. The web based preparing capacity of Continual BERT empowers versatile learning on new information streaming in a period consecutive way, particularly fitting to the mind-boggling measure of COVID-19 writing distributed consistently.

NIR: Neural Index Run, BioBERT-NLI model [18], RFRR: Relevance criticism with BERT-based re-positioning benchmark, Hybrid list with both an upset and neural file for positioning. Neural positioning is useful, yet has a few downsides, which might be eased when combined with a conventional reversed list. "Self-supervised context-aware Covid-19 document exploration through atlas grounding" published by Dusan Grujicic et al., addresses How to learn venture sentences into an actual space characterized by a three-dimensional anatomical chart book? Utilizing Bidirectional Encoder Representations from Transformers – BERT [19], Text-to-map book planning objective, 1) It permits us to imagine clinical content in truly important space, discovering groups of reports coordinated by life structures 2) It permits us to look for and re-trieve text by exploring through a physical space.3) There is a factual favorable position to demonstrating clinical content in the 3D space as anatomically related foundations will in general be near each other.

"Interactive Visualization and Simplified Pattern Discovery in the COVID-19 Open Research Dataset(CORD-19)" published by Wuraola Fisayo Oyewusi focuses on Step by step instructions to picture a huge dataset and how to consider the examples, relationship dependent on frequencies [20] of illness and compound elements utilizing Scispacy, Scispacy Named Entity Recognition Model (en ner bc5cdr md) prepared on the Bio Creative V Chemical-Disease Relations (BC5CDR) where Each of the information point in the content perception [22], when clicked shows the content information in various settings from the CORD-19 dataset.

2.2 SUMMARY OF LITERATURE SURVEY

- Most of the work carried out focused on repurposing drugs for curing cancer or different types of cancers.
- Previous work reported used the reviews and posts or drug review post about drugs from social media.
- The work done primarily employed a pattern dependant combination extraction method to extract ill gene and gene ill combination.
- Graph based techniques using ABC model for protein sequences and genomic sequences where used.

2.3 RESEARCH GAPS AND CHALLENGES

- Techniques used was tedious in terms of identifying the sequences of genes and protein, and also the output required expert analysis.
- Most of the work so far reported where on analysis of drug post reviews, tweets and not on articles or literature of covid-19.
- Pretrained drug dataset on articles is not available instead reviews and tweets dataset about drugs are available.
- Previously reported work so far focused only on the sentiment of one drug at a time, and also used a platform called natural language understanding

CHAPTER 3

3.1 SOFTWARE AND HARDWARE REQUIREMENTS

This chapter gives the details about the software and hardware requirements needed for the project execution.

Software Requirements required for the project

Serial No.	Requirement	Specification
1	PLATFORM	Python Notebook, Jupyter.
2	INSTALLATIONS	Sentiment Analysis Libraries, File Formatting Libraries.
3	PACKAGES	NLP Sentiment and Deep Learning
4	OPERATING SYSTEM	WINDOWS 10

Table 1. Software Requirements

Hard requirements required for the project

Serial No.	Requirement	Specification
1	Processor used	AMD Radeon Processor
2	Memory	16 GB Random Access Memory
3	Disk Storage	200 GB

Table 2. Hardware Requirements

CHAPTER 4

4.1 PROPOSED ARCHITECTURE

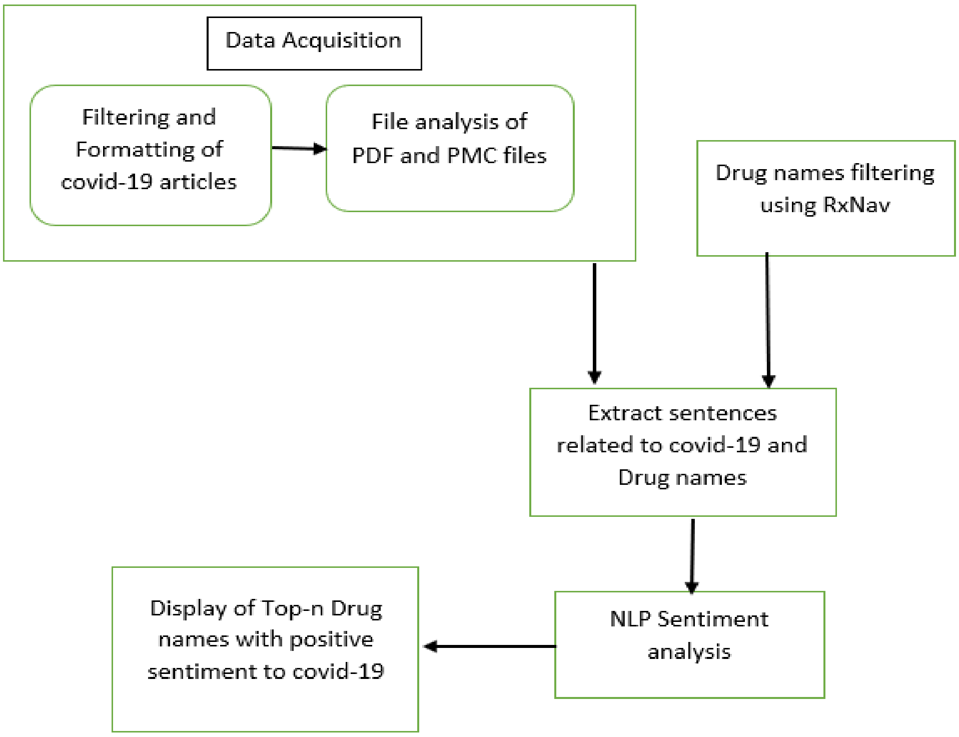

Data Acquisition
Filtering and Formatting of covid-19 articles
File analysis of PDF and PMC files
Drug names filtering using RxNav
Extract sentences related to covid-19 and Drug names
NLP Sentiment analysis
Display of Top-n Drug names with positive sentiment to covid-19

Fig. 4.1 Diagram of Proposed Architecture

Fig 4.1 refers to the proposed system architecture, which consists of many steps like data acquisition, filtering and formatting the articles and drug names. Sentences containing covid-19 and drug names was filtered and was further used for processing. Later, NLP sentiment analysis was employed to find the drug names of most relevant drugs. The following Fig. 4.2 will describe the used steps more clearly.

4.2 LOW LEVEL DIAGRAM OF PROPOSED ARCHITECTURE

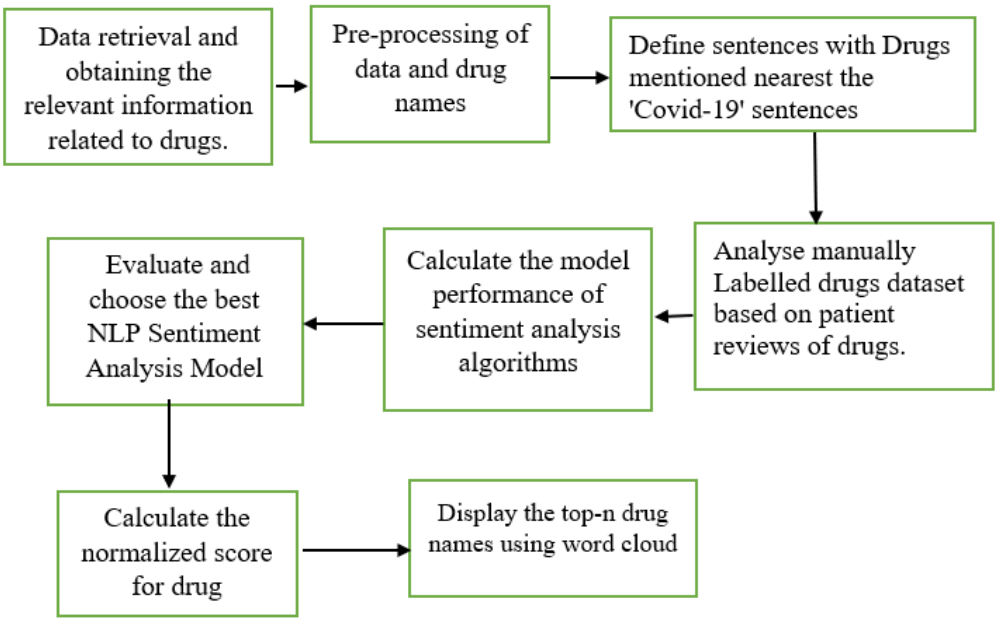
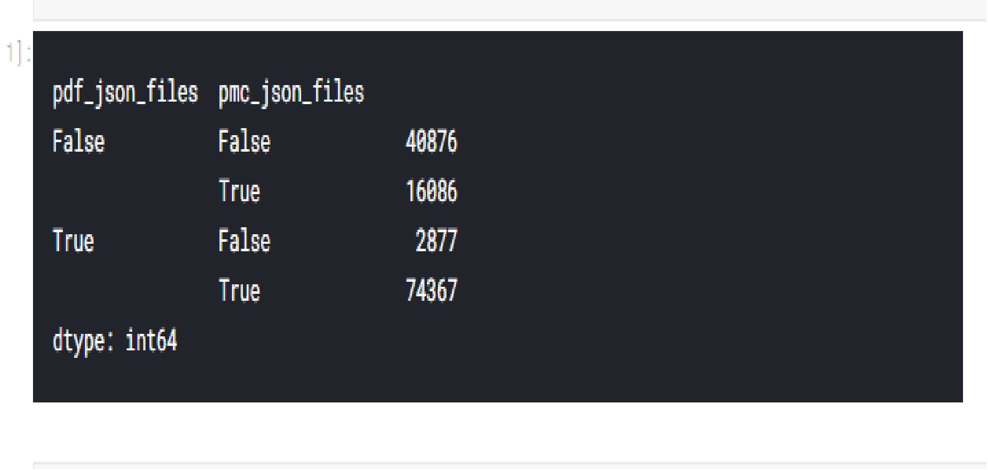

Data retrieval and obtaining the relevant information related to drugs.
Pre-processing of data and drug names
Define sentences with Drugs mentioned nearest the 'Covid-19' sentences
Analyse manually Labelled drugs dataset based on patient reviews of drugs.
Calculate the model performance of sentiment analysis algorithms
Evaluate and choose the best NLP Sentiment Analysis Model
Calculate the normalized score for drug
Display the top-n drug names using word cloud

Fig. 4.2 Low level Diagram of Proposed Architecture

Step 1- Data Acquisition

The dataset was vast and included literature/articles other than corona virus, the articles needed to be filtered and formatted. Filtered articles with file links - pdf json files or pmc json files as seen in Fig. 4.3 – from the year 2020.



```
pdf_json_files  pmc_json_files
False           False           40876
                True            16086
True            False            2877
                True            74367
dtype: int64
```

Fig. 4.3 Number of PMC and PDF files

Cord-19 documents sorted by publication date equals 19785 documents for the year 2020. The files contained lots of links attached to it, which needed a preliminary analysis of the file before pre-processing. The majority of the articles were published in the month of April according to the analysis as seen in Fig. 4.4.

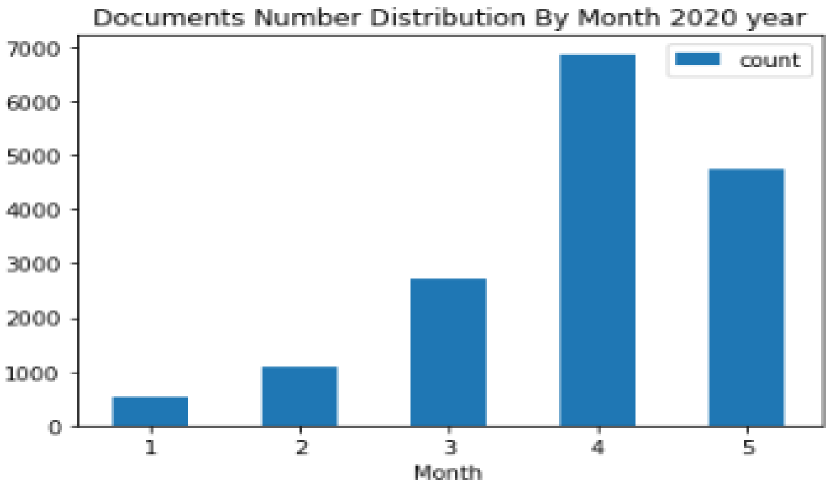
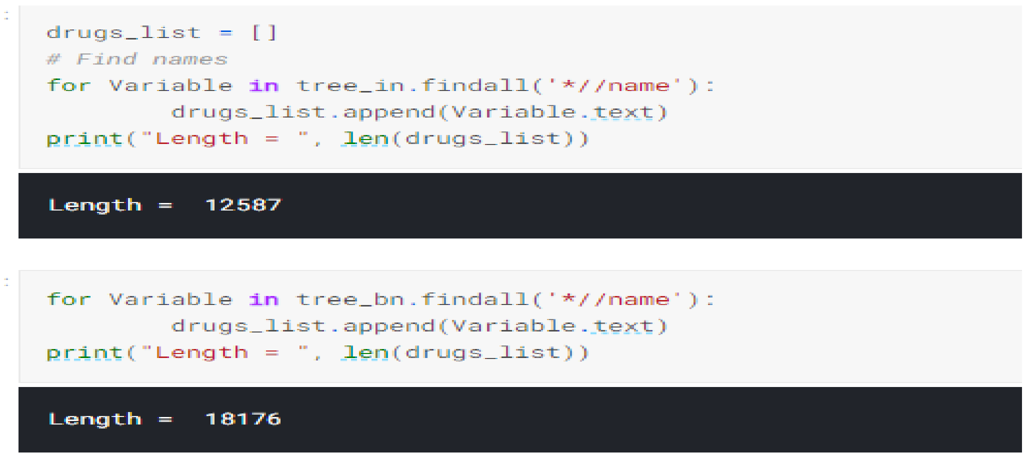


Fig. 4.4 Graph that shows distribution of articles published in 2020

After the filtering of articles, the focused turned to the filtering of drug names for which RxNav API was used. Drug names were extracted from RxNav, which included 18176 unique entries.



```
drugs_list = []
# Find names
for Variable in tree_in.findall('*//name'):
        drugs_list.append(Variable.text)
print("Length = ", len(drugs_list))
```

```
Length =  12587
```

```
for Variable in tree_bn.findall('*//name'):
        drugs_list.append(Variable.text)
print("Length = ", len(drugs_list))
```

```
Length =  18176
```

Fig. 4.5 Drug list from RxNav

Fig.4.5 shows the number of drug names RxNav consisted. RxNav is a drug information browser that includes RxNorm and RxTerms, among others. RxNav uses the names and codes in its constituent vocabularies to find medications in RxNorm. For both branded and generic therapeutic medications, RxNav provides connections to active compounds, medication components, and related brand names. Additionally, RxNav provides code lists and links to Daily Med package inserts. RxNav may get the RxTerms record for a medicine as well as clinical data from MED-RT, such as pharmacologic classifications, modes of action, and physiologic effects.

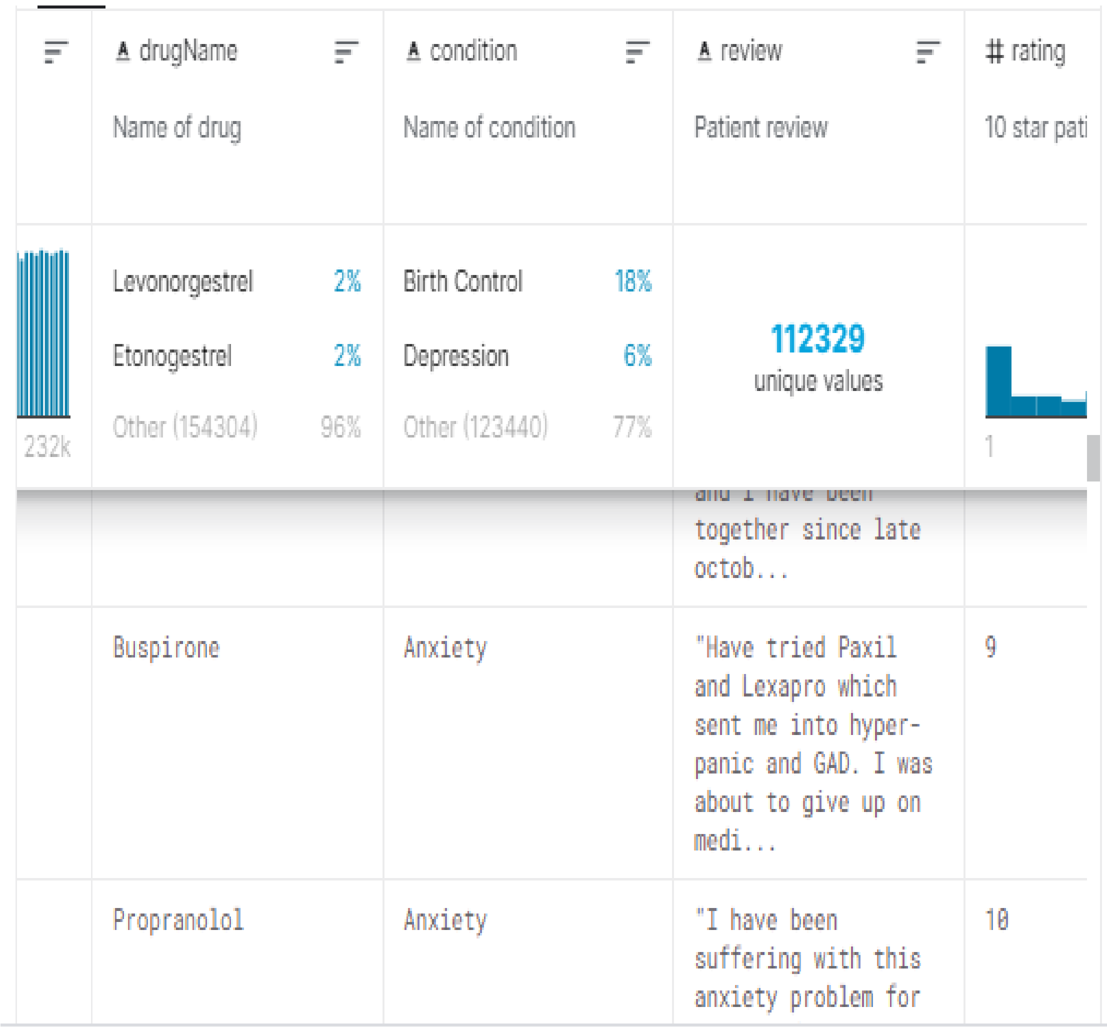
Step 2- Extracting sentences related to Drugs and covid-19

The sentences with covid-19 and Drugs names with a distance of two sentences between them are retrieved, and statistics regarding the distribution of Drugs names in each text are calculated. Then employ a natural language processing (NLP) sentiment analysis model, the difficulty is that all existing sentiment models are pre-trained on social media datasets (tweets, movie/hotel reviews).

Step 3- NLP Sentiment Analysis

Natural Language Processing is a discipline that brings together linguistics and computer technology. The goal is for machines to comprehend or "understand" natural language in order to perform tasks that humans do, such as translating languages and answering questions. In lexicon-based approaches, a feeling is determined by its semantic direction and the strength of each word in the phrase. This involves the employment of a pre-defined lexicon that divides words into positive and negative categories. In this project, vader sentiment analysis and text blob analysis was used to find the top drug names related to covid-19 with positive sentiments.

A group of words is frequently used to express a text message. After assigning individual scores to all of the words, the final emotion is calculated using a pooling algorithm.



	drugName Name of drug	condition Name of condition	review Patient review	# rating 10 star pati
232k	Levonorgestrel 2% Etonogestrel 2% Other (154304) 96%	Birth Control 18% Depression 6% Other (123440) 77%	112329 unique values	1
			and I have been together since late octob...	
	Buspirone	Anxiety	"Have tried Paxil and Lexapro which sent me into hyper-panic and GAD. I was about to give up on medi...	9
	Propranolol	Anxiety	"I have been suffering with this anxiety problem for	10

Fig. 4.6 Sentiment Labelled Dataset for Drug reviews

Sentiment labelled dataset for drug reviews as shown in Fig. 4.6 was used in order to train the model for sentiment analysis. The dataset had ratings between 1-10 for drug reviews given by patients, I used thresholding strategies to convert these ratings into categorical classes like neutral, positive and negative.

.

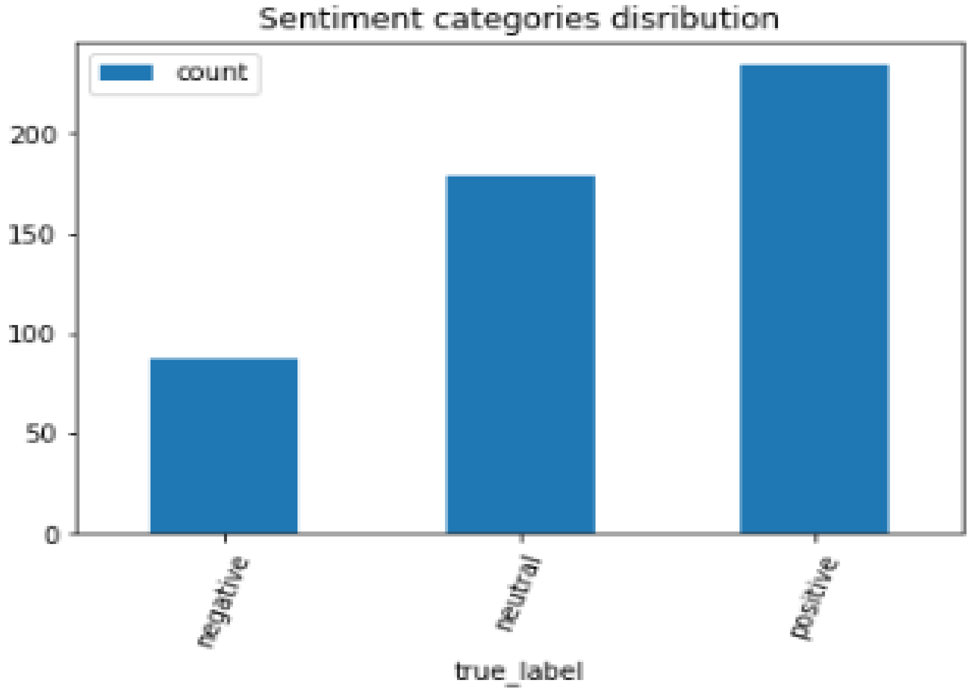


Fig. 4.7 Graph that shows sentiment category distribution

As seen in Fig. 4.7, The distribution of this dataset has many positive categories when compared to other categories. Once the analysis of training data was ready, it was fed to vader sentiment analyser which was customized for this purpose.

Vader is a sentiment known lexicon and a sentiment analysis tool based on rules. Vader is designed for social media data and can produce impressive results when combined with Twitter data., Facebook, etc. Released on May 22, 2020. The majority of Vader's ratings were obtained from Amazon's Mechanical Turk. VADER keeps (and even improves on) the advantages of classic sentiment lexicons like LIWC: it's bigger, yet it's just as easy to analyse, understand, apply (without substantial learning/training), and extend.

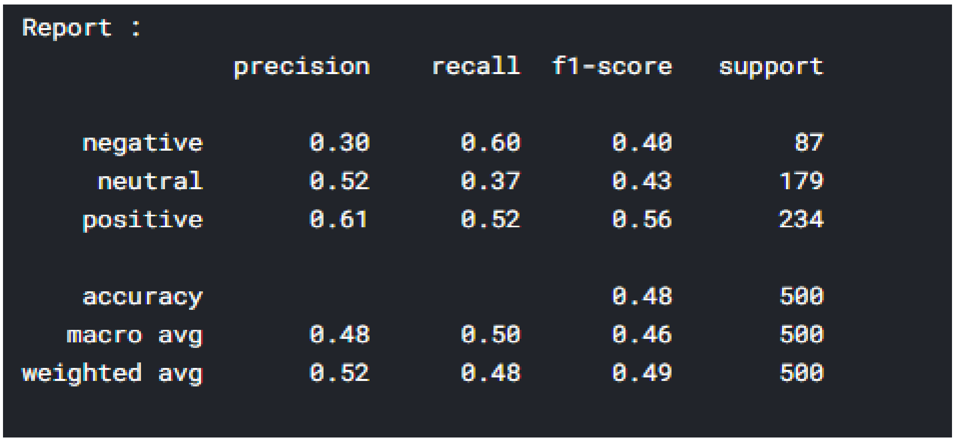
The VADER sentiment lexicon, like LIWC (but unlike certain other lexicons or machine learning models), is of gold-standard quality and has been human-validated. VADER differs from LIWC in that it is more responsive to social media sentiment expressions while also generalising more favourably to other domains.

The two types of sentiment words are base type and comparison type emotion words. All of the words in the preceding instances are of the base type. Comparative sentiment words (which include superlative sentiment words) are used to convey comparative and superlative opinions. Emotion detection is a text analysis tool that looks for signs of different emotional states. To figure out what is what and why, lexicons and machine learning algorithms are commonly used. It appears to be a routine extraction of the specific understanding on the surface. However, to really grasp the gist of the emotion extraction, some heavy lifting algorithms are required.

Context helps even the most well-intentioned emotion mining company. While a human can obtain context with little effort, the algorithm has a very different perspective. Algorithms are unable to predict what actions they will need to take in order to achieve the optimal results. They must be set in order to achieve the greatest results. As a result, a component that manages the message's context must be included in the sentiment analysis model. Because it depicts the links between words in a text as well as their relationships in terms of parts of speech, text vectorization is crucial.

When compared to complex machine learning algorithms, however, VADER's simplicity has significant advantages. To begin with, it is both fast and computationally efficient without sacrificing precision. A corpus that takes a fraction of a second to analyse with VADER can take hours or tens of minutes to analyse with more complex models like SVM (if training is required) or minutes if the model has been previously trained when running directly from a standard modern laptop computer with typical, moderate specifications (e.g., 3GHz processor and 6GB RAM). VADER makes the inner workings of the sentiment analysis engine more available (and hence more interpretable) to a broader human audience outside the computer by exposing both the vocabulary and the rule-based model.

The first step is to tokenize the sentences, which breaks them up into multiple words. The sentiment value of each word sentence can be easily estimated by comparing internal sentiment lexicon. Despite the absence of machine learning, this library parses each tokenized word, compares it to its lexicon, and returns polarity scores.



```
Report :
              precision    recall  f1-score   support

    negative       0.30      0.60      0.40        87
     neutral       0.52      0.37      0.43       179
    positive       0.61      0.52      0.56       234

    accuracy                           0.48       500
   macro avg       0.48      0.50      0.46       500
weighted avg       0.52      0.48      0.49       500
```

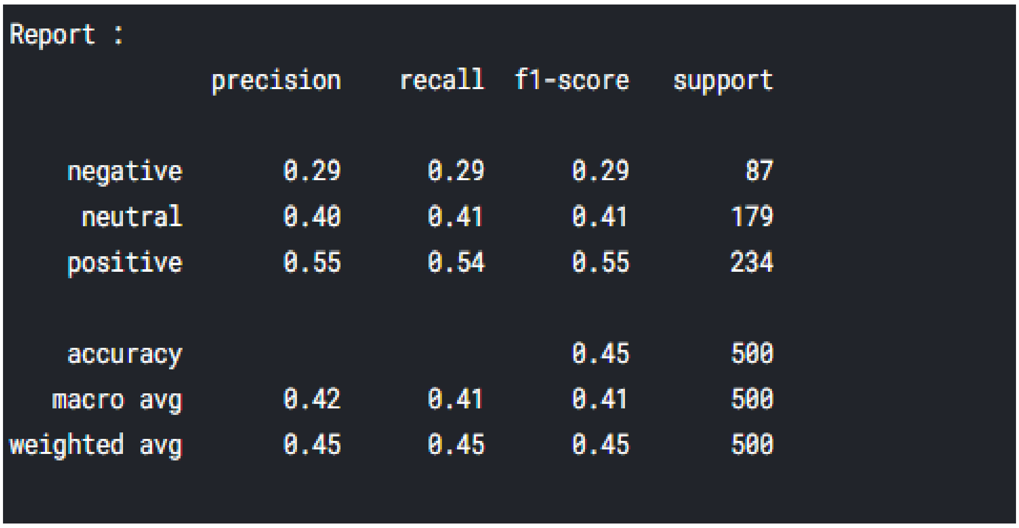
Fig. 4.8 Report of Vader Sentiment Analysis

Fig. 4.8 shows the report obtained when vader sentiment analysis was used, which showed a weighted average precision of 0.52, and recall of 0.48. This generates a sentiment score for the entire input. VADER is a free, open-source Python library that can be installed via pip. It doesn't require any training data, and it's fast enough to handle data that's available in real time.

Vader has the ability to function in a variety of environments. Since Vader was trained on online data, it was not able to capture the sentiments within the articles. While the Vader model isn't completely correct, it is very rudimentary. The Vader model revealed that while it is not that accurate, it is quite representative.

Vader has the ability to function in many domains. Vader does not require any foreign data in order to comprehend the feeling of a text that includes emotions, slang terms, and much more. Vader is a master of social media text. As a result, Vader is an accurate tool for coming up with new ideas for online writing. When vader gave a low accuracy even though it worked good for other domains, I used text blob sentiment analysis which consists of sentiment lexicon a list of predefined words to give scores to each word, which are then meaned out using a weighted average to provide an overall sentence sentiment score.

The Text blob sentiments module, an NLTK classifier trained on a movie reviews corpus, includes two sentiment analysis implementations: Pattern Recognizer and NB Recognizer. The default implementation is Pattern Recognizer, but you can change it by passing another implementation to a Text Blob's function Object () [native code]. It can be time consuming to pass taggers, sentiment analysers, and tokenizers to a large number of Text Blobs. To keep your code tidy, we may utilise the Blobber object to create Text Blobs that give the same models.



```
Report :
              precision    recall  f1-score   support

    negative       0.29      0.29      0.29        87
     neutral       0.40      0.41      0.41       179
    positive       0.55      0.54      0.55       234

    accuracy                           0.45       500
   macro avg       0.42      0.41      0.41       500
weighted avg       0.45      0.45      0.45       500
```

Fig. 4.9. Report of Text Blob Sentiment Analysis

As shown in Fig. 4.9, all samples contribute equally to the final averaged measure when it is micro-averaged. All classes contribute equally to the final averaged metric when it is macro-averaged. Weighted-averaged: the contribution of each class to the overall average is weighted by its size. Confusion matrix classifies algorithm's performance. If there are unbalanced number of observations in each class or if dataset has more than two classes, classification accuracy alone can be misleading. In practise, when we strive to improve our model's precision, the recall suffers, and vice versa. The F1-score encapsulates both trends.

Text Blob employed the Natural Language Toolkit to fulfil its duties. NLTK is a library that enables users to deal with categorization, classification, and a variety of other tasks after giving separate scores to each word. The final sentiment is derived using some form of pooling operation, such as taking an average of the individual scores. Text Blob is a simple library that allows for complex textual data analysis and operations. Each word receives three scores: "polarity," "subjectivity," and "intensity." We are only interested in "polarity": positive vs. negative (-1.0 => +1.0)

Deep learning has recently been presented as a solution to certain NLP tasks. Its key benefit is that, unlike supervised learning, it does not necessitate the manual tuning of features based on expert knowledge and available linguistic resources. Deep learning (DL) is regarded as the next step in the evolution of machine learning, which is promising for sentiment analysis as shown in Fig 4.10.

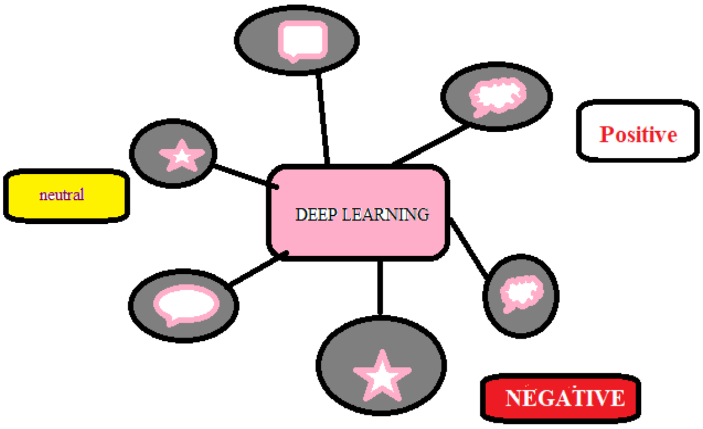


Fig. 4.10 Sentiment Analysis using Deep Learning

Despite their complexity, traditional Machine Learning algorithms offer the advantage of being machine-like. They necessitate a considerable lot of topic knowledge as well as human aid that can only perform what it was designed to do. In this aspect, deep learning holds more promise for artificial intelligence creators and the remaining of the world. To put it another way, a deep learning strategy is a technique that allows you to learn more profoundly. Using its hidden layer architecture, learn categories in phases, starting with low-level categories like letters, moving up to somewhat higher-level categories like words, and finally to higher level categories like sentences. It is also known as an artificial neural network since it connects algorithms to imitate how the human brain operates. Sentiment analysis models can be trained to grasp text beyond simple definitions, read for context, sarcasm, etc., and understand the writer's true mood and feeling by using the power of deep learning.

A deep learning model's first phase training takes a long time and often requires millions of data points before it can grasp on its own. Deep learning models can split words, paragraphs, and entire publications into individual opinion units using natural language processing (NLP). Deep Learning surpasses traditional techniques when the amount of data is large. When dealing with little amounts of data, however, traditional automated learning techniques are preferable. Deep learning techniques demand high end technology to train in a reasonable length of time. When there is a lack of domain understanding for attribute introspection, Deep Learning techniques outperform others since feature engineering is not an issue.

The model could perform subject classification to categorise each statement into predetermined categories once it was broken down into opinion units. You can't give machine algorithms language and expect them to solve your problems because they can only interpret and process numeric input (especially floating-point data). Instead, you'll have to alter your data numerous times before it takes on a representative numeric shape. A distributed representation of the input layers is encoded by the hidden layers in a DNN. Word embeddings are dimensional space representations of words that are distributed representations of words.

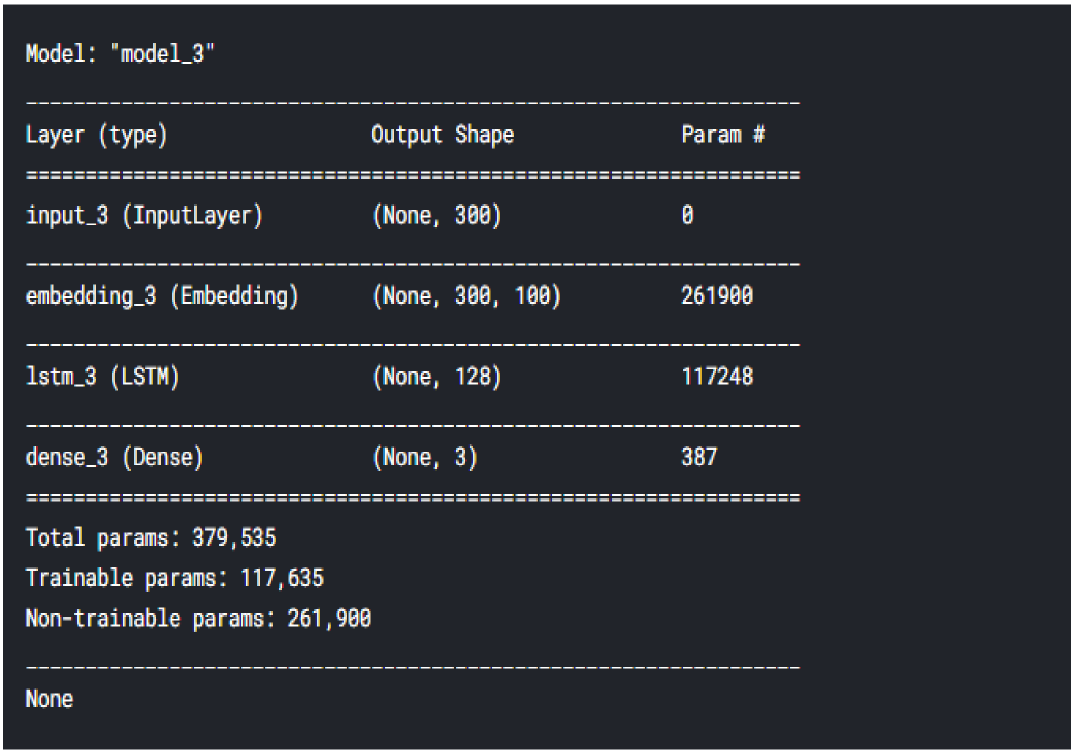
They are often the top hidden units in a deep learning architecture that has been trained on a substantial quantity of data. These pretrained embeddings can be fed into another deep neural architecture designed for a specific job. Convolution filters are efficient at processing data in a matrix or grid representation, but they only capture sequential and linear patterns in a narrow area and can miss long-range connections between variables.

The LSTM was created to address this issue. It's a sort of recurrent neural network that makes use of LSTM units. Each LSTM unit uses four information gates to decide which new and existing data to add to or remove from the information flow when processing sequential data. The hyperparameters were trained in most of the trials by using the development set or cross validation were given in the corpus. Word embedding techniques like Glove and Word2Vec have shown to be particularly effective in converting words into dense vectors.

The vector is small, and none of the indexes in it is a number. Converting words into their appropriate numeric indexes is the initial stage in word embeddings. Because sentences might be different lengths, the Tokenizer class returns sequences with varied lengths as well. The sequence will have a maximum length of 300 characters.

The remaining indices will be padded with zeros for sentences that are less than 300 characters long. The remaining indices will be shortened for sentences longer than 300 characters. Changing the labels from continuous feelings to - 0,1,2 is required. With dense vectors, deep learning models converge faster than with sparse ones.

In this project, I created a model that included one input layer (embedding layer), one LSTM layer with 128 units, and one dense layer that also served as an output layer. Because there are three possible outputs, the number of neurons will be three, and the activation function will be a categorical function like SoftMax.



```
Model: "model_3"
-----------------------------------------------------------------
Layer (type)                 Output Shape              Param #
=================================================================
input_3 (InputLayer)         (None, 300)               0
-----------------------------------------------------------------
embedding_3 (Embedding)      (None, 300, 100)          261900
-----------------------------------------------------------------
lstm_3 (LSTM)                (None, 128)               117248
-----------------------------------------------------------------
dense_3 (Dense)              (None, 3)                 387
=================================================================
Total params: 379,535
Trainable params: 117,635
Non-trainable params: 261,900
-----------------------------------------------------------------
None
```

Fig. 4.11. LSTM model Description

Fig. 4.11 refers to the description about LSTM model, which describes the input layer shape, embedding layer shape, trainable parameters, non-trainable parameters and total parameters. The accuracy obtained using deep learning algorithm was much better than other sentiment analysis such as text blob analysis and vader sentiment analysis. The outcome is acceptable because the true and projected outcomes are almost identical. RNNs are a good choice for processing sequential data, however they have a short-term memory problem. The gating mechanism controls the flow of information in RNNs, which helps to solve the problem.

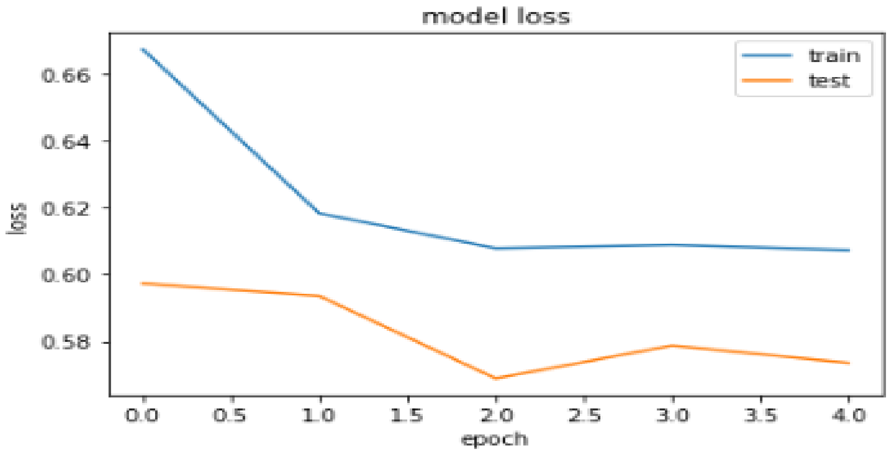
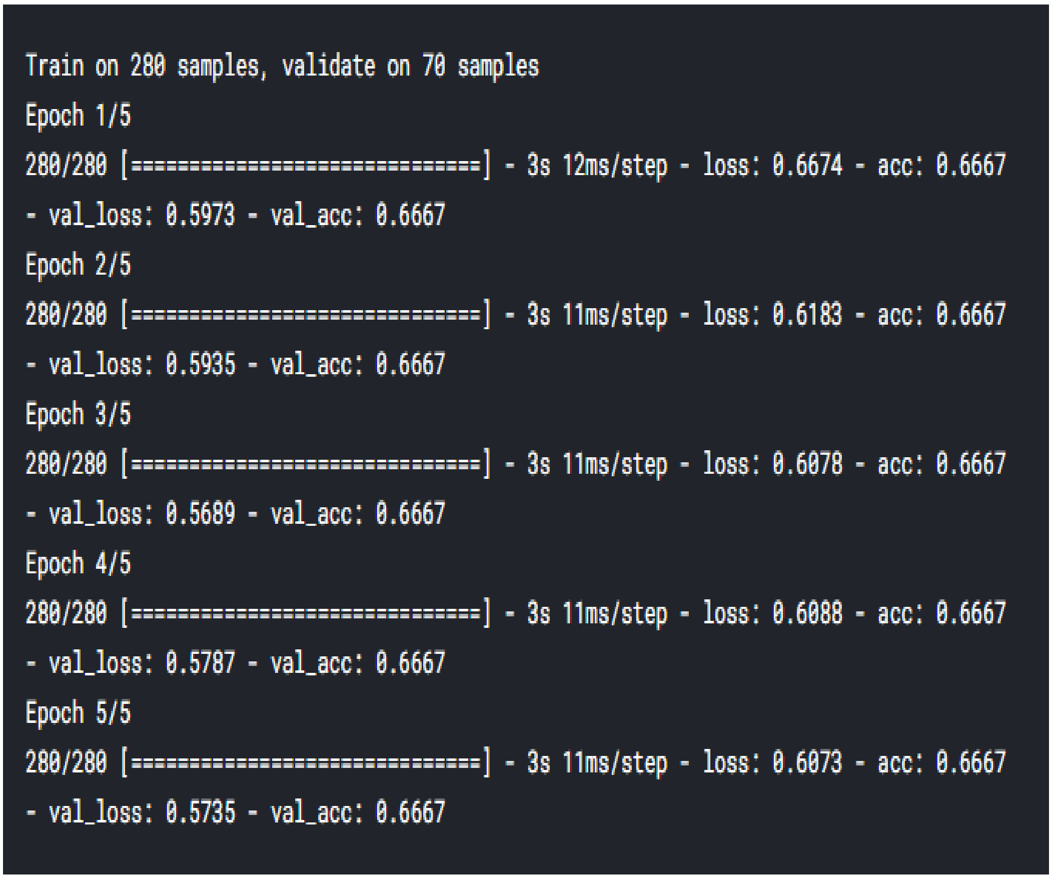


Fig. 4.12. Graph Representing Model Loss

As we seen in Fig. 4.12, we can limit the number of epochs to two because after that accuracy stays the same. The number of epochs determines how frequently the weights of the network are changed.



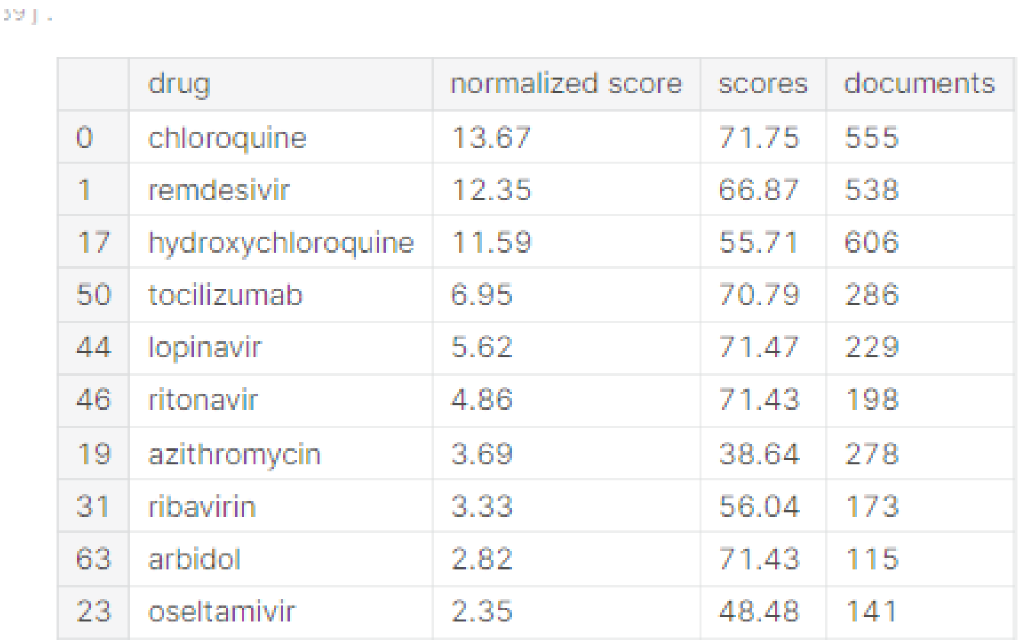
```
Train on 280 samples, validate on 70 samples
Epoch 1/5
280/280 [==============================] - 3s 12ms/step - loss: 0.6674 - acc: 0.6667
- val_loss: 0.5973 - val_acc: 0.6667
Epoch 2/5
280/280 [==============================] - 3s 11ms/step - loss: 0.6183 - acc: 0.6667
- val_loss: 0.5935 - val_acc: 0.6667
Epoch 3/5
280/280 [==============================] - 3s 11ms/step - loss: 0.6078 - acc: 0.6667
- val_loss: 0.5689 - val_acc: 0.6667
Epoch 4/5
280/280 [==============================] - 3s 11ms/step - loss: 0.6088 - acc: 0.6667
- val_loss: 0.5787 - val_acc: 0.6667
Epoch 5/5
280/280 [==============================] - 3s 11ms/step - loss: 0.6073 - acc: 0.6667
- val_loss: 0.5735 - val_acc: 0.6667
```

Fig. 4.13. Optimal number of epochs

Fig. 4.13 shows that validation loss did not change much after the second epoch, also the validation accuracy did not change after the second epoch. The neural network's weights are changed the same number of times as the number of epochs increases, and the boundary line shifts from underfitting to optimal to overfitting. Steps per epoch denotes the number of batches to be chosen for each epoch. If 500 steps are chosen, the network will train for 500 batches in order to complete one epoch. In this experiment 280 batches were chosen. The number of times an algorithm visits a data set is defined as an epoch. We don’t witness a change in the performance of the model, so we restrict it to two epochs.

Step 4- Normalizing the Drug score.

Normalization is a technique often applied as part of data preparation for machine learning. The goal of normalization is to change the values of numeric columns in the dataset to use a common scale, without distorting differences in the ranges of values or losing information. This was done to project the obtained result to the word cloud,



	drug	normalized score	scores	documents
0	chloroquine	13.67	71.75	555
1	remdesivir	12.35	66.87	538
17	hydroxychloroquine	11.59	55.71	606
50	tocilizumab	6.95	70.79	286
44	lopinavir	5.62	71.47	229
46	ritonavir	4.86	71.43	198
19	azithromycin	3.69	38.64	278
31	ribavirin	3.33	56.04	173
63	arbidol	2.82	71.43	115
23	oseltamivir	2.35	48.48	141

Fig. 4.14. Results showing Normalized Drug Scores

Fig. 4.14 shows the normalized drug score which is defined as, Normalized sentiment score: (Average Score) * (Number of Document with drug mentioned / Total Number of Documents). Chloroquine had the highest number of documents which was around 555, and Remdesivir has 538 supporting documents when normalized. The least number of documents were accounted to Arbidol.

Step 5- Displaying the output obtained

Further to bring all the drug scores to same scale, normalization of drug scores was carried out. The final result was displayed using a word cloud on NLP platform. Word clouds, also known as tag clouds, are graphical representations of word frequency that give words that appear more frequently in a source text more emphasis. The larger the term in the image, the more frequently it appeared in the document. Word clouds are a simple and cost-effective way to visualise text data. One of the difficulties in deciphering word clouds is that the display highlights the frequency of words rather than their significance.

CHAPTER 5

5.1 IMPLEMETATION DETAILS

This chapter includes the implantation details of the project, or the experimental setup required for the project. Python unreal environment allows to instal Python packages in a dedicated area for a specific application rather than installing them worldwide. Python libraries are essential for developing applications in machine learning, data processing, and more. Libraries like pandas, json, nltk, sklearn, copy, keras, word cloud, sentiment analyser were imported, also various installation dependencies also were included in order to support the library.

Hence the corpus consisted of various articles from vivid domain, filtering was required in order to gather the articles related to corona virus alone. The articles obtained was formatted differently which when inputted to algorithm will be a cumbersome task for the machine. So, formatting of articles was required to have a smooth processing of articles.

There were around 16085 pdf files in the articles and around 74367 pmc json files , which was obtained for further processing Total count of articles was more than a lakh which contributed to the larger dataset. There were many APIs that provided drug names for the repurposing but the most popular among them was RxNav which had more than thousands of drug names repository which also contained the ingredients of the drugs.

There were around 18k drug names in the RxNav APIs website, in which this project aims to extract the drug names related to corona virus with positive notion. The API contained other information of the drug like manufacture’s name, the brand name and more.

Pre-processing was carried out to a great extent because there was many unwanted information embedded within the text information, and also complete text of the articles was used other than metadata alone. Since the articles was large in size many NLP sentiment algorithms was carried out to find which one among them will give the correct emotions with respect to drugs being repositioned. When traditional library fed algorithm failed to give the output, sentiment analysis using deep learning was tried.

5.2 TESTING

Many factors influence sentiment accuracy, including the sort of data you're working with and the people who hand-tagged your sentiment library. Human analysts tend to agree on a decent score when analysing the attitude (positive, negative, or neutral) of a text piece, according to studies.

However, if you are using natural language processing to automate sentiment analysis, you will want to be sure.

NLP allows for the construction of test cases using Natural Language, which eliminates the need to learn or understand a set of rules. When test cases are written in Natural Language, they are not only readable but also understandable by users of various skill levels.

5.3 RESULTS AND DISCUSSION

After reviewing the various ways that computational drug repositioning strategies and models have been used to identify new therapeutic interactions, we can conclude that each strategy and approach has its own set of advantages and disadvantages, and that combining vivid strategies and approaches often results in a higher success and accuracy rate.

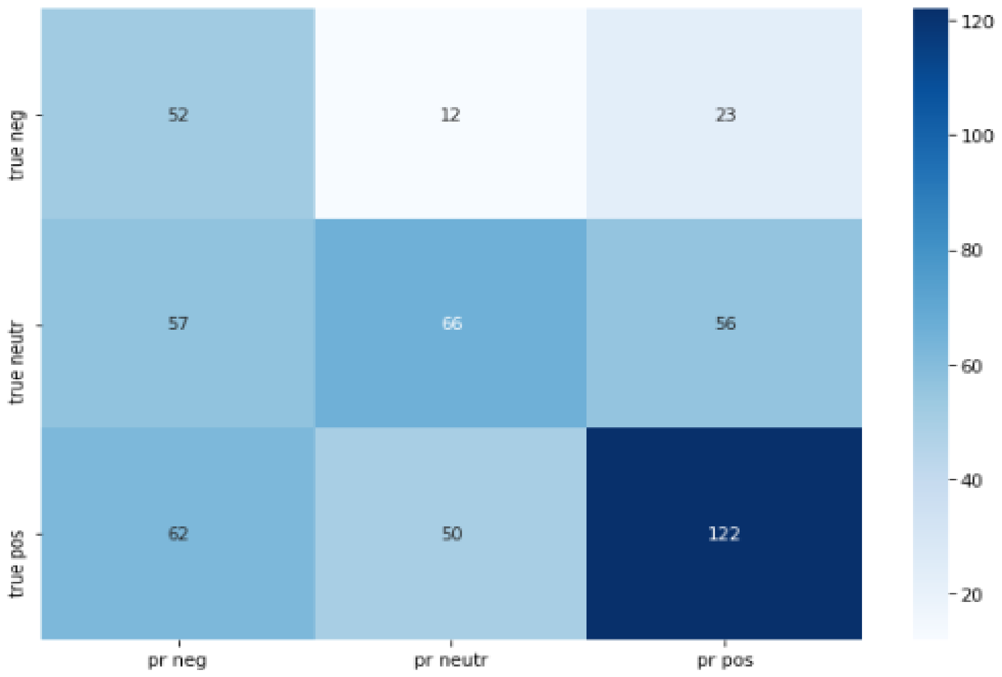
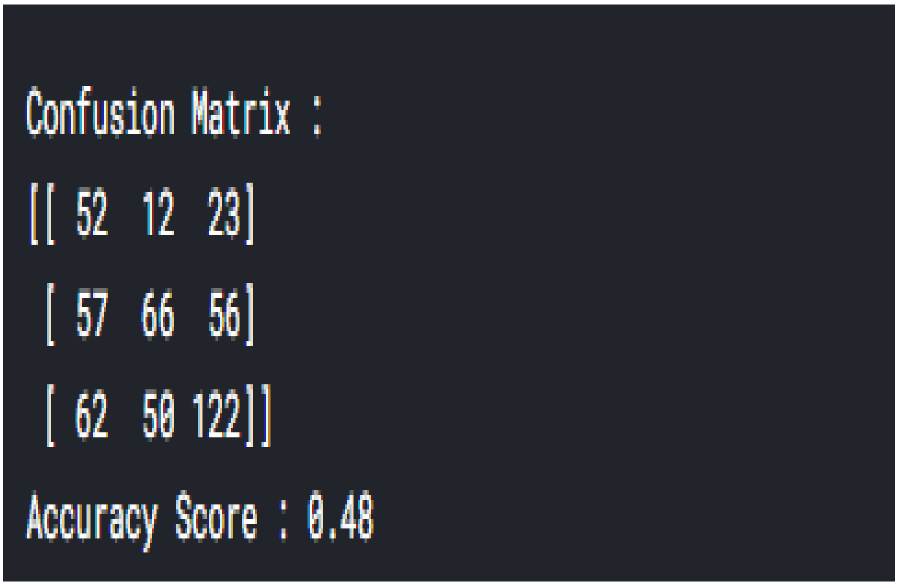


Fig. 5.1. Confusion matrix of Vader sentiment analysis

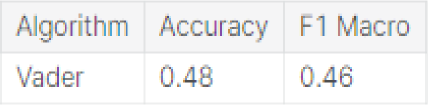
The output obtained using vader sentiment analysis was not satisfactory as seen in Fig 5.1, the confusion matrix shows that there are many miss classifications. Since it is tuned for social media content, it performs best on the content you can find on social media. However, it still offers acceptable F1 Scores on other test sets, and provides a comparable performance compared to complex statistical models



```
Confusion Matrix :
[[ 52  12  23]
 [ 57  66  56]
 [ 62  50 122]]
Accuracy Score : 0.48
```

Fig. 5.2 Accuracy obtained using Vader sentiment analysis

As shown in Fig. 5.2 the accuracy obtained using vader sentiment analysis was just 48 percentage, which cannot be considered as a good accuracy for a real time data.



Algorithm	Accuracy	F1 Macro
Vader	0.48	0.46

Fig 5.3 Summary table of Vader sentiment analysis

In Fig. 5.3 it is clear that F1 Macro was 0.46 which was lesser that accuracy obtained, macro F1-score (short for macro-averaged F1 score) is a metric for evaluating the quality of problems that include numerous binary labels or classes.

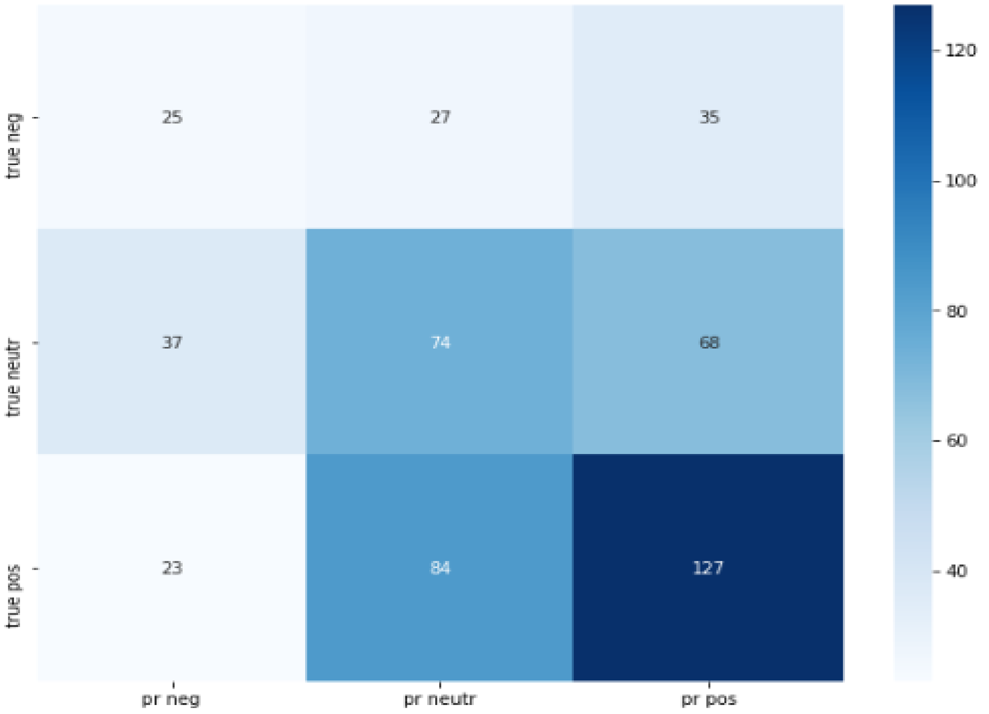
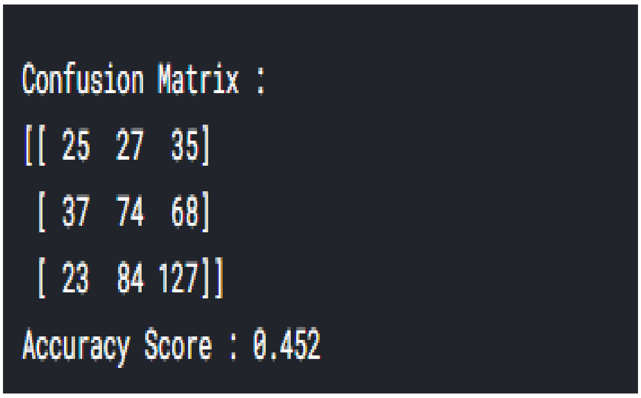


Fig. 5.4. Confusion matrix of Text Blob sentiment analysis

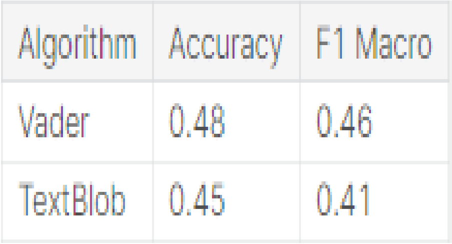
Fig. 5.4 shows the confusion matrix obtained when text blob sentiment analysis was used. It neither showed a great accuracy, in fact showed an accuracy lesser than vader sentiment analysis. It lacks features like as dependency parsing, word vectors, and so on Fig. 5.5.



```
Confusion Matrix :
[[ 25  27  35]
 [ 37  74  68]
 [ 23  84 127]]
Accuracy Score : 0.452
```

Fig. 5.5. Accuracy obtained using blob text sentiment analysis

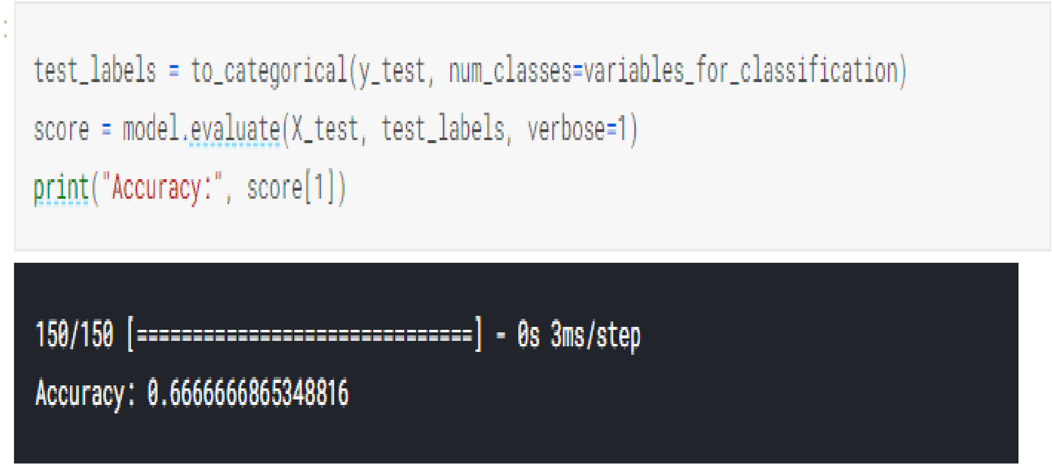
As seen, Fig 5.5 shows the summary table obtained using both vader sentiment analysis and text blob sentiment analysis. When compared to text blob sentiment analysis, vader performed better by giving a better accuracy and F1 Macro



Algorithm	Accuracy	F1 Macro
Vader	0.48	0.46
TextBlob	0.45	0.41

Fig. 5.6. Summary table of Vader sentiment analysis and Text Blob sentiment analysis

.



```
test_labels = to_categorical(y_test, num_classes=variables_for_classification)
score = model.evaluate(X_test, test_labels, verbose=1)
print("Accuracy:", score[1])
```

```
150/150 [==============================] - 0s 3ms/step
Accuracy: 0.6666666865348816
```

Fig. 5.7. Results obtained using Deep Learning Sentiment Analysis

With around 67 percent accuracy rate as seen in Fig. 5.7 , the Top-3 Drugs with the greatest favourable feelings normalised scores were identified. Chloroquine was retrieved from 555 documents with a score of 13.67. Remdesivir retrieved from 538 papers with a score of 12.35. Hydroxy chloroquine was found in 606 papers with a score of 11.59. There was no improvement after hyper parameter tuning.

The number of epochs did not help the accuracy much either. For deep learning sentiment analysis 150 batch size was enough as it had tensor mechanism to incorporate the processing.

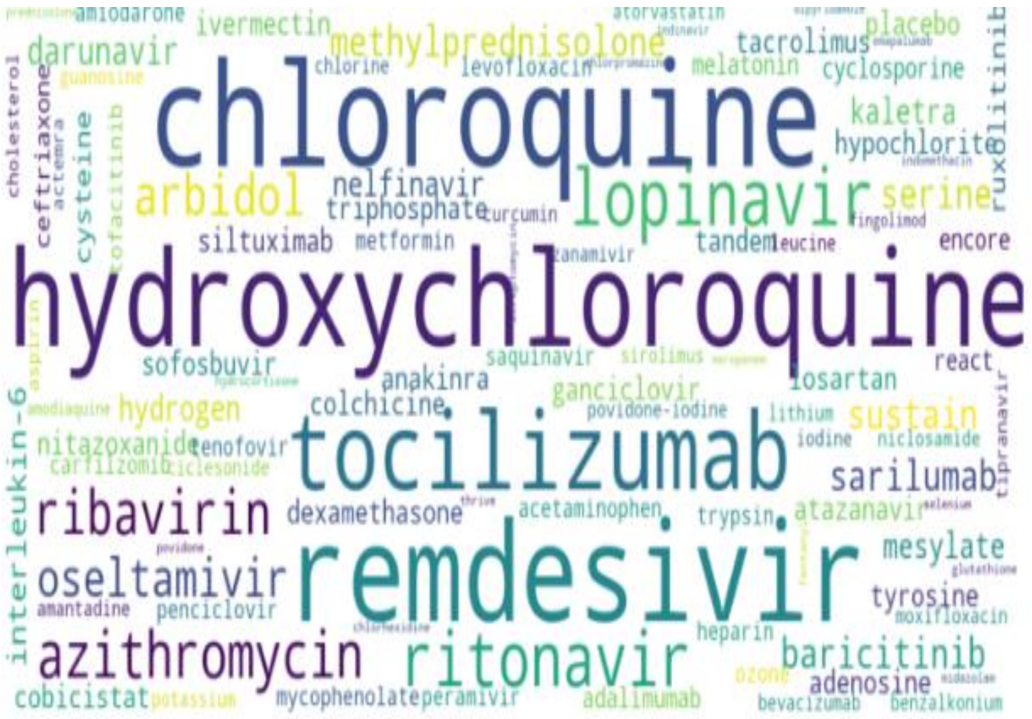


Fig. 5.8. Word Cloud Displaying Top-n Drug Names

We can see from the chart Fig. 5.8 above that hydroxychloroquine and chloroquine are the most positively mentioned medications, although there are a few other drugs like Remdesivir that are commonly utilised. As a result, the most important drugs for this investigation are hydroxychloroquine and chloroquine.

A word cloud is a basic but effective visual representation object for text processing that displays the most frequently used words in larger, bolder letters and with varied colours. The lesser the importance of a term, the smaller it is

Another group with scores and references that are 2-3 times lower. Tocilizumab was discovered in 286 documents with a score of 6.95. Lopinavir was discovered in 229 papers with a score of 5.62. Ritonavir was detected in 198 papers with a score of 4.86. Azithromycin was discovered in 278 papers with a score of 3.69. Ribavirin was detected in 173 papers with a score of 3.33. Arbidol was detected in 115 papers with a score of 2.82.

The only medicine licenced by the FDA to treat coronavirus illness is Remdesivir (COVID-19). Remdesivir is also being researched in conjunction with other drugs. Several hydroxychloroquine studies received a lot of attention in the early months of the epidemic, especially in the media.

Many clinical trials are presently underway to investigate various potential COVID-19 treatments, such as monoclonal antibodies. Older drugs (usually used to treat different disorders) are also being tested to see if they are beneficial against COVID-19. Various recommendations have been made about the use of remdesivir. Due to a lack of data, the National Institutes of Health (NIH) supports using remdesivir for select COVID-19 hospitalised patients, but the World Health Organization (WHO) does not suggest it for any COVID-19 patients. Remdesivir is also being researched in conjunction with other drugs. Not all remdesivir studies have yielded positive results. A comprehensive, randomised research published by the World Health Organization (WHO) on October 15, 2020 indicated that remdesivir had little or no influence on death rates in hospitalised COVID-19 patients. Regardless of whether patients received remdesivir or not, the death rate was around 11%. These findings have not yet been peer-reviewed.

Antiviral medicine oseltamivir is used to treat influenza (flu). The findings of a hospital in Wuhan, China, were not encouraging. Tamiflu was given to 124 of the 138 hospitalised patients, along with additional treatments. 85 patients (62%) were remained hospitalised at the end of the trial, and 6 had died.

In vitro research, studies done in a petri dish or test tube rather than in animals or humans. SARS-CoV-2, the virus that causes COVID-19, is resistant to both hydroxychloroquine and chloroquine. These drugs acted by interfering with the chemical environment of human cell membranes in these experiments.

The virus was unable to enter and proliferate within the cells as a result of this. A medication's ability to act in vitro does not always imply that it will work in the human body. Researchers hastened to explore the effects of hydroxychloroquine and chloroquine in hospitalised COVID-19 patients based on these in vitro findings. Early results were made public, prompting hospitals all around the world to begin employing these drugs.

Tocilizumab reduced the likelihood of progression to the composite endpoint of mechanical ventilation or death in hospitalised patients with Covid-19 pneumonia who were not receiving mechanical ventilator, but it did not enhance survival. Although a few small studies suggested that they might be useful for treating COVID-19 in hospitalised patients with mild instances, national health agencies (such as the FDA) now agree that hydroxychloroquine and chloroquine do not function for preventing or treating COVID-19.

The National Institutes of Health advises against using them for COVID-19. Antibiotics like azithromycin are routinely used to treat bacterial illnesses including bronchitis and pneumonia. It has been proven to have some in vitro activity against viruses such as influenza A and Zika, but not against the MERS-causing coronavirus.

Tocilizumab is an IL-6 inhibitor that has been licenced for use in the treatment of rheumatoid arthritis and juvenile idiopathic arthritis. (Both of these conditions are inflammatory.) It acts by inhibiting the protein interleukin-6 (IL-6), which is involved in our normal immunological responses. IL-6 is a cytokine (signalling protein) that usually signals other cells to activate the immune system, but too much activation might be problematic. A cytokine storm, a potentially catastrophic problem in which the immune system goes wild and inflammation spirals out of control, is one possible serious issue with an overactive immune system. People who are infected with COVID-19 may experience cytokine storms as their immune systems ramp up to battle the virus. Tocilizumab helps to quiet down the immune system by suppressing IL-6, and it's also thought to aid with cytokine storms.

Tocilizumab research began with a French trial that found that persons who received tocilizumab were less likely to require ventilation or die. Another Italian trial indicated that individuals who received tocilizumab had a decreased death rate, while both groups used ventilators at about the same rate. Tocilizumab, on the other hand, had no effect on COVID-19 individuals with early-stage pneumonia. Tocilizumab did not benefit hospitalised COVID-19 patients with severe pneumonia, according to a phase 3 research conducted by the manufacturer.

COVID-19 is also being studied with Kevzara (sarilumab), a drug that operates similarly to tocilizumab. The initial results were not encouraging. Patients with severe symptoms who received Kevzara fared worse than those who received placebo, whereas patients with even more severe (critical) symptoms fared better.

One study looked at azithromycin in combination with hydroxychloroquine as a COVID-19 treatment. They reported that 93% of patients had cleared the infection after 8 days, but there was no control group, so we don't know if people would have cleared the virus on their own if the drugs hadn't been used..

Kaletra is an HIV treatment that combines lopinavir and ritonavir, two antiviral drugs. These antiviral medicines may have some effect against SARS and MERS, according to in vitro and clinical trials of individuals who have previously taken them (infections caused by other coronaviruses). There is a scarcity of data about Kaletra's use in COVID-19.

Casirivimab and Imdevimab are two monoclonal antibodies that make up this antibody cocktail. Imdevimab and Casirivimab are monoclonal antibodies to human immunoglobulin G-1 (IgG1) that act against the SARS-CoV-2 spike protein (the virus that causes COVID-19).

The antibody mixture prevents the virus from adhering to the human cell and entering it. The cocktail antibody, which has been approved for emergency use authorization (EUA), is reported to have the ability to treat mild-to-moderate Covid-19 infection in adults and children over the age of 12 who are at a higher risk of developing severe disease and do not require oxygen. When compared to seronegative individuals who only received standard care, the monoclonal antibody treatment reduced the primary outcome of 28-day death in seronegative patients by one-fifth. This suggests that there would be six fewer deaths for every 100 patients treated with the antibody mixture.

CHAPTER 6

6.1 CONCLUSION

To sum up, the fundamental document analysis was carried out in order to find Top Drugs Names with positive attitudes in relation to Covid-19 treatment. During the analysis, I purified and analysed the CORD-19 dataset, identifying and visualising some statistical properties. Chloroquine, remdesivir, and hydroxychloroquine were shown to be the top three medications for covid-19 with the highest favourable context.

Combinational drugs have been shown to be useful in treating covid-19 and reducing the severity of the condition, as well as hospitalisation. In patients with mild to moderate Covid-19 symptoms, the combinational therapy is believed to lower the likelihood of hospitalisation by 70%. Patients who got a combination of baricitinib and remdesivir recovered 1 day faster than those who just received remdesivir.

The goal of this effort was to identify drug names that are likely to aid against COVID-19 based on favourable features or merging two or more drugs mentioned about the medicine in medical documents retrieved from various sources such as PubMed, WHO, and others.

6.2 CHALLENGES FACED

Despite the fact that we have some excellent computational drug repositioning or repurposing models, establishing robust models is still a difficult endeavour. The difficulty of putting theoretical computational ideas into practise is one of the most significant obstacles, due to the difficulty of translating such theoretical concepts to imitate the behaviour of living organisms, as well as other challenges such as missing, skewed, or erroneous data. Another problem with computational drug repositioning or repurposing models is the lack of credible gold-standard datasets to test their efficacy.

Researchers can either develop their own gold-standard dataset and compare and evaluate their suggested models using generally used evaluation methods, or they can partition their input into training, testing, and validating sets and apply K-fold cross validation. Despite the challenges of computational drug repositioning research, sentiment analysis based on deep learning will surpass traditional algorithms.

Because the medical field is so complex, a person without a medical degree may find it difficult to comprehend the meaning of some statements. Medical publications should not use pre-trained models based on social media data (tweets, movie/hotel reviews). The NTLK sentence parser isn't very good for this medical corpus. The drug names were limited to those available in the RxNav Rest API repository.

6.3 FUTURE WORK

- Add a weight to the score depending on the section's name or parse Drugs Names from the special document's sections alone.

- Improve Sentiment Analysis using Machine Learning and Deep Learning Models.

- Improve the splitting and detection of sentences.

- At the very least, find or develop Labelled Sentiment datasets for Medical Domain linked Drugs Treatment.

REFERENCES

1. Oscar Araque, Kyriaki Kalimeri, Lorenzo Guti "Sentiment and Moral Narratives during COVID-19", ACL 2020 Workshop NLP-COVID conference.

2. Ramya Tekumalla Juan M Banda, "Characterization of Potential Drug Treatments for COVID-19 using Social Media Data and Machine Learning ", arXiv preprint arXiv:1903.10676.

3. Hyeju Jang, Emily Rempel, Giuseppe Carenini1 , Naveed Janjua, " Exploratory Analysis of COVID-19 Related Tweets in North America to Inform Public Health Institutes", Bioinformatics 36, no. 4 (2020): 1234-1240.

4. Anna Kruspe, Matthias Haberle ," Cross-language sentiment analysis of European Twitter messages during the COVID-19 pandemic", ACL 2020 Workshop NLP-COVID conference.

5. Hsih-Te Yang, Jiun-Huang Ju, "Literature-based discovery of new candidates for drug repurposing", Briefings in Bioinformatics 2018.

6. Gaocai Dong, Ping Zhang, Jingya Yang, Dongdong Zhang, Jing Peng, "A Systematic Framework for Drug Repurposing based on Literature Mining" IEEE International Conference on Bioinformatics and Biomedicine 2019.

7. Mohamed Abdel Basset , Hossam Hawash Mohamed Elhoseny, Ripon K. Chakrabortty and Michael Ryan "DeepH-DTA: Deep Learning for Predicting Drug-Target Interactions: A Case Study of COVID-19 Drug Repurposing", IEEE International Conference 2020.

8. Boshu Ru, Dingcheng Li, Lixia Yao, "Detecting Serendipitous Drug Usage in Social Media with Deep Neural Network Models", IEEE International Conference on Bioinformatics and Biomedicine 2018.

9. Seyedeh Shaghayegh Sadeghi , Mohammad Reza Keyvanpour, “An Analytical Review of Computational Drug Repurposing”, IEEE International Conference 2020.

10. Guangsheng Wu, Juan Liu, Caihua Wang, “Semi-supervised graph cut algorithm for drug repositioning by integrating drug, disease and genomic associations”, IEEE International Conference on Bioinformatics and Biomedicine 2018.

11. Yadi Zhou, Fei Wang, Jian Tang, Ruth Nussinov, Feixiong Cheng, “FNN to classify drugs into pharmaceutical therapeutic classes based on the drugs”, The Lancet Digital Health journal 2020.

12. Christine Herlihy and Yuelin Liu, “Automated Task-Informed Document Retrieval on the COVID-19 Open Research Dataset Using Topic Modeling”. ACL 2020 Submission.

13. Cécile Robin, Georgeta Bordea, Bianca Pereira, John Philip McCrae, Paul Buitelaar, “A Term Extraction Approach to Expert Finding on the COVID-19 Open Research Dataset”. ACL Conference, 2020.

14. Oscar William Lithgow-Serrano, Alejandra Lopez-Fuentes, Yalbi Balderas-Martinez, Fabio Rinaldi., “A smart literature exploration environment for COVID-19 literature”. In European conference on information retrieval, 2020.

15. Matteo Muffo, Aldo Cocco, Mattia Messina, Enrico Bertino, “Question Answering tool for COVID-19”. ACL Conference, 2020.

16. Wuraola Fisayo Oyewusi, “Interactive Visualization and Simplified Pattern Discovery in the COVID-19 Open Research Dataset (CORD-19)”. ACL Conference, 2020.

17. Dusan Grujicic, Gorjan Radevski, Tinne Tuytelaars, and Matthew B. Blaschko, “Self-supervised context-aware Covid-19 document exploration through atlas grounding”. MedRix journal, 2020.

18. Bernal Jimenez Guti ´ errez, Juncheng Zeng, Dongdong Zhang, Ping Zhang, Yu Su, “Document Classification for COVID-19 Literature”. medRxiv and bioRxiv journals, 2020.

19. Adam Poliak, Max Fleming, Cash Costello, Kenton Murray, Shivani Pandya, “Collecting Verified COVID-19 Question Answer Pairs”. ACL conference, 2020.

20. Jong Won Park, “Continual BERT: Continual Learning for Adaptive Extractive Summarization of COVID-19 Literature”. In Asian conference on information retrieval, 2020.

21. Vincent Nguyen, Maciej Rybinski1, Sarvnaz Karimi Zhenchang Xing, “Searching Scientific Literature for Answers on COVID-19 Questions”. ACL conference, 2020.