MACHINE LEARNING BASED APPROACHES FOR DETECTING COVID-19 USING CLINICAL TEXT DATA

S Pooja Redy[1],Devasoth Anusha[2], B Aravind Reddy[3],Jarpula Srikanth[4],
Ch Pavan Kumar Reddy[5]

[2,3,4,5] UG Scholars, Department of CSE, **AVN Institute of Engineering and Technology,**Hyderabad, Telangana, India.

[1]Assistant Professor, Department of CSE, **AVN Institute of Engineering and Technology**, Hyderabad, Telangana, India.

Abstract :

Technology advancements have a rapid effect on every field of life, be it medical field or any other field. Artificial intelligence has shown the promising results in health care through its decision making by analysing the data. COVID-19 has affected more than 100 countries in a matter of no time. People all over the world are vulnerable to its consequences in future. It is imperative to develop a control system that will detect the coronavirus. One of the solution to control the current havoc can be the diagnosis of disease with the help of various AI tools. In this paper, we classified textual clinical reports into four classes by using classical and ensemble machine learning algorithms. Feature engineering was performed using techniques like Term frequency/inverse document frequency (TF/IDF), Bag of words (BOW) and report length. These features were supplied to traditional and ensemble machine learning classifiers. Logistic regression and Multinomial Naïve Bayes showed better results than other ML algorithms by having 96.2% testing accuracy. In future recurrent neural network can be used for better accuracy.

INTRODUCTION

In December 2019, the novel coronavirus appeared in the Wuhan city of China [1] and was reported to the World Health Organization (W.H.O) on 31st December 2019. The virus created a global threat and was named as COVID-19 by W.H.O on 11th February 2020 [1]. The COVID-19 is the family of viruses including SARS, ARDS. W.H.O declared this outbreak as a public health emergency [2] and mentioned the following; the virus is being transmitted via the respiratory tract when a healthy person comes in contact with the infected person. The virus may transmit between persons through other roots which are currently unclear. The infected person shows symptoms within 2–14 days, depending on the incubation period of the middle east respiratory syndrome (MERS), and the severe acute respiratory syndrome (SARS). According to W.H.O the signs and symptoms of mild to moderate cases are dry cough, fatigue and fever while as in severe cases dyspnea (shortness of breath), Fever and tiredness may occur [3, 4]. The persons having other diseases like asthma, diabetes, and heart disease are more vulnerable to the

virus and may become severely ill. The person is diagnoses based on symptoms and his travel history. Vital signs are being observed keenly of the client having symptoms. No specific treatment has been discovered as on 10th April 2020, and patients are being treated symptomatically. The drugs like hydroxychloriquine, antipyretic, anti-virals are used for the symptomatic treatment. Currently, no such vaccine is developed for preventing this deadly disease, and we may take some precautions to prevent this disease. By washing hands regularly with soap for 20 s and avoiding close contact with others by keeping the distance of about 1 m may reduce the chances of getting affected by this virus. While sneezing, Covering the mouth and nose with the help of disposable tissue and avoiding the contact with the nose, ear and mouth can help in its prevention. SARS is an airborne disease that appeared in 2003 in China and affected 26 countries by having 8 K cases in the same year and transferred from person to person. The signs and symptoms of SARS are fever, cold, diarrhoea, shivering, malaise, myalgia and dyspnea. The ARDS (acute respiratory distress syndrome) is characterized by rapid onset of inflammation in lungs which leads to respiratory failure and its signs and symptoms are bluish skin colour, fatigue and shortness of breath. ARDS is diagnosed by PaO2/FiO2 ratio of less than 300 mm Hg. Till 10th of April 2020, almost 1.6 million confirmed cases of coronavirus are detected around the globe. Almost 97 K persons have died and 364 K persons have recovered from this deadly virus [5]. Figure 1 shows the worldwide data regarding coronavirus. Since no drug or vaccine is made

for curing the COVID-19. Various paramedical companies have claimed of developing a vaccine for this virus. Less testing has also given rise to this disease as we lack the medical resources due to pandemic. Since thousands and thousands are being tested positive day by day around the globe, it is not possible to test all the persons who show symptoms.

Apart from clinical procedures, machine learning provides a lot of support in identifying the disease with the help of image and textual data. Machine learning can be used for the identification of novel coronavirus. It can also forecast the nature of the virus across the globe. However, machine learning requires a huge amount of data for classifying or predicting diseases. Supervised machine learning algorithms need annotated data for classifying the text or image into different categories. From the past decade, a huge amount of progress is being made in this area for resolving some critical projects. Recent pandemic has attracted many researchers around the globe to solve
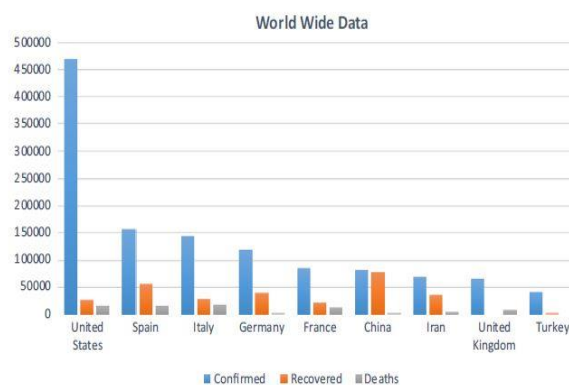
this problem. Data provided by John Hopkins University in the form of X-ray images and various researchers build a model of machine

learning that classifies X-ray image into COVID-19 or not. Since the latest data published by Johns Hopkins gives the metadata of these images. The data consists of clinical reports in the form of text in this paper, we are classifying that text into four different categories of diseases such that it can help in detecting coronavirus from earlier clinical symptoms. We used supervised machine learning techniques for classifying the text into four different categories COVID, SARS, ARDS and Both (COVID, ARDS). We are also using ensemble learning techniques for classification. Section 2 gives the literature survey regarding the proposed work. The framework for detecting coronavirus from clinical text data is being discussed in Sects. 3 and 4 gives the experimental results of the proposed framework and Sect. 5 concludes our work.

1.1. OBJECTIVE

The main objective of this project is It is imperative to develop a control system that will detect the coronavirus. One of the solution to control the current havoc can be the diagnosis of disease with the help of various AI tools.

1.2. PROBLEM SPECIFICATION

Technology advancements have a rapid effect on every field of life, be it medical field or any other field. Artificial intelligence has shown the promising results in health care through its decision making by analysing the data. COVID-19 has affected more than 100 countries in a matter of no time. People all over the world are vulnerable to its consequences in future. It is imperative to develop a control system that will detect the coronavirus. One of the solution to control the current havoc can be the diagnosis of disease with the help of various AI tools

1.3. METHODOLOGIES

Machine learning and natural language processing use big data-based models for pattern recognition, explanation, and prediction. NLP has gained much interest in recent years, mostly in the field of text analytics, Classification is one of the major task in text mining and can be performed using different algorithms [6]. Kumar et al. [7] performed a SWOT analysis of various supervised and unsupervised text classification algorithms for mining the unstructured data. The various applications of text classification are sentiment analysis, fraud detection, and spam detection etc. Opinion mining is majorly being used for elections, advertisement, business etc. Verma et al. [8] analysed Sentiments of Indian government projects with the help of the lexicon-based dictionary. The machine learning has changed the perspective of diagnosis by giving great results to diseases like diabetes and epilepsy. Chakraborti et al. [9] detected epilepsy using machine learning approaches, electroencephalogram (EEG) signals

are used for detecting normal and epileptic conditions using artificial neural networks (ANN). Sarwar et al. [10] diagnosis diabetes using machine learning and ensemble learning techniques result indicated that ensemble technique assured accuracy of 98.60%. These purposes can be beneficial to diagnose and predict COVID-19. Firm and exact diagnosis of COVID-19 can save millions of lives and can produce a massive amount of data on which a machine learning (ML) models can be trained. ML may provide useful input in this regard, in particular in making diagnoses based on clinical text, radiography Images etc. According to Bullock et al. [11], Machine learning and deep learning can replace humans by giving an accurate diagnosis. The perfect diagnosis can save radiologists' time and can be cost-effective than standard tests for COVID-19. X-rays and computed tomography (CT) scans can be used for training the machine learning model. Several initiatives are underway in this regard. Wang and Wong [12] developed COVID-Net, which is a deep convolutional neural network, which can diagnose COVID-19 from chest radiography images. Once the COVID-19 is detected in a person, the question is whether and how intensively that person will be affected. Not all COVID-19 positive patients will need rigorous attention. Being able to prognosis who will be affected more severely can help in directing assistance and planning medical resource allocation and utilization. Yan et al. [13] used machine learning to develop a prognostic prediction algorithm to predict the mortality risk of a person that has been infected, using data from (only) 29 patients at Tongji Hospital in Wuhan, China. Jiang et al. [14] proposed a machine learning model that can predict a person affected with COVID-19 and has the possibility to develop acute respiratory distress syndrome (ARDS). The proposed model resulted in 80% of accuracy. The samples of 53 patients were used for training their model and are restricted to two Chinese hospitals. ML can be used to diagnose COVID-19 which needs a lot of research effort but is not yet widely operational. Since less work is being done on diagnosis and predicting using text, we used machine learning and ensemble learning models to classify the clinical reports into four categories of viruses.

1.4. CONTRIBUTIONS

In this project we are using traditional and classical machine learning algorithms to predict COVID-19 disease. In traditional algorithms we are using Logistic Regression, Naïve Bayes, SVM and Decision Tree and in classical algorithms we are using Bagging, AdaBoost, Random Forest and Stochastic Gradient Boosting classifier. In all algorithms Logistic Regression giving better performance.

1.4. LAYOUT THESIS

Data collection

As W.H.O declared Coronavirus pandemic as Health Emergency. The researchers and hospitals give open access to the data regarding this pandemic. We have collected from an open-source data repository GitHub.1 In which about 212 patients data is stored which have shown symptoms of corona virus and other viruses. Data consists of about 24 attributes namely patient id, offset, sex, age, finding, survival, intubated, went_icu, needed_supplemental_ O2, extubated, temperature, pO2_saturation, leukocyte_ count, neutrophil count, lymphocyte count, view, modality, date, location, folder, filename, DOI, URL. License. Clinical notes and other notes.

Relevant dataset

Since our work is regarding text mining so we extracted clinical notes and findings. Clinical notes consist of text while as the attribute finding consist label of the corresponding text. About 212 reports were used and their length was calculated. We consider only those reports that are written in the English language. Figure 3 gives the length distribution of clinical reports that are written in English. The clinical reports are labelled to their corresponding classes. In our dataset, we have four classes COVID, ARDS, SARS and Both (COVID, ARDS). Figure 4 shows the different classes in which clinical text is being categorized and corresponding report length.

Preprocessing

The text is unstructured so it needed to be refined such that machine learning can be done. Various steps are being followed in this phase; the text is being cleaned by removing unnecessary text. Punctuation and lemmatisation are being done such that the data is refined in a better way. Stopwords, symbols, Url's, links are removed such that classification can be achieved with better accuracy. Figure 5 shows the main steps in preprocessing.

2. LITERATURE SURVEY

Title: **new coronavirus associated with human respiratory disease in china**

Author: Z.-K. Zhang, M. C. Y. Cho, C.-W. Wang, C.-W. Hsu, C.-K. Chen, and S. Shieh.

Emerging infectious diseases, such as severe acute respiratory syndrome (SARS) and Zika virus disease, present a major threat to public health1,2,3. Despite intense research efforts, how, when and where new diseases appear are still a source of considerable uncertainty. A severe respiratory disease was recently reported in Wuhan, Hubei province, China. As of 25 January 2020, at least 1,975 cases had been reported since the first patient was hospitalized on 12 December 2019. Epidemiological investigations have suggested that the outbreak was associated with a seafood market in Wuhan. Here we study a single patient who was a worker at the market and who was admitted to the Central Hospital of Wuhan on 26 December 2019 while experiencing a severe respiratory syndrome that included fever, dizziness and a cough. Metagenomic RNA sequencing4 of a sample of

bronchoalveolar lavage fluid from the patient identified a new RNA virus strain from the family Coronaviridae, which is designated here 'WH-Human 1' coronavirus (and has also been referred to as '2019-nCoV'). Phylogenetic analysis of the complete viral genome (29,903 nucleotides) revealed that the virus was most closely related (89.1% nucleotide similarity) to a group of SARS-like coronaviruses (genus Betacoronavirus, subgenus Sarbecovirus) that had previously been found in bats in China5. This outbreak highlights the ongoing ability of viral spill-over from animals to cause severe disease in humans.

Title: Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study.

Author: A. Dorri, S. S. Kanhere, R. Jurdak, and P. Gauravaram,

Background: In December, 2019, a pneumonia associated with the 2019 novel coronavirus (2019-nCoV) emerged in Wuhan, China. We aimed to further clarify the epidemiological and clinical characteristics of 2019-nCoV pneumonia. Methods: In this retrospective, single-centre study, we included all confirmed cases of 2019-nCoV in Wuhan Jinyintan Hospital from Jan 1 to Jan 20, 2020. Cases were confirmed by real-time RT-PCR and were analysed for epidemiological, demographic, clinical, and radiological features and laboratory data. Outcomes were followed up until Jan 25, 2020.

Findings: Of the 99 patients with 2019-nCoV pneumonia, 49 (49%) had a history of exposure to the Huanan seafood market. The average age of the patients was 55·5 years (SD 13·1), including 67 men and 32 women. 2019-nCoV was detected in all patients by real-time RT-PCR. 50 (51%) patients had chronic diseases. Patients had clinical manifestations of fever (82 [83%] patients), cough (81 [82%] patients), shortness of breath (31 [31%] patients), muscle ache (11 [11%] patients), confusion (nine [9%] patients), headache (eight [8%] patients), sore throat (five [5%] patients), rhinorrhoea (four [4%] patients), chest pain (two [2%] patients), diarrhoea (two [2%] patients), and nausea and vomiting (one [1%] patient). According to imaging examination, 74 (75%) patients showed bilateral pneumonia, 14 (14%) patients showed multiple mottling and ground-glass opacity, and one (1%) patient had pneumothorax. 17 (17%) patients developed acute respiratory distress syndrome and, among them, 11 (11%) patients worsened in a short period of time and died of multiple organ failure. Interpretation: The 2019-nCoV infection was of clustering onset, is more likely to affect older males with comorbidities, and can result in severe and even fatal respiratory diseases such as acute respiratory distress syndrome. In general, characteristics of patients who died were in line with the MuLBSTA score, an early warning model for predicting mortality in viral pneumonia. Further investigation is

needed to explore the applicability of the MuLBSTA score in predicting the risk of mortality in 2019-nCoV infection.

Title: Text classification algorithms for mining unstructured data: a SWOT analysis

Author: E. Bertino and N. Islam,

It has become increasingly crucial and imperative to facilitate knowledge extraction for decision support and deliver targeted information to analysts that span wide application domains. Interestingly, the buzzing term "big data" which is estimated to be 90% unstructured further makes it difficult to tap and analyze information with traditional tools. Text mining entails defining a process which transforms and substitutes this unstructured data into a structured one to discover knowledge. Use of classification algorithms to intelligently mine text has been studied extensively across literature. This study predominantly surveys the text classification algorithms employed in the process of mining unstructured data to report a conclusive analysis on the trend of their use in terms of their respective strengths, weaknesses, opportunities and threats (SWOT). The scope of these algorithms is then explored apropos the application area of sentiment analysis, a typical text classification task. A mapping which determines the unexplored social media technologies and the extent of use of these algorithms within respective social media is proffered to give an insight to the amount of work that has been done

in the domain of machine learning based sentiment analysis on social media.

Title: **Diagnosis of diabetes type-II using hybrid machine learning based ensemble model**

Author: C. Zhang and R. Green.

The work done in this paper exhibits an expert system based ensemble model in diagnosing type-II diabetes. Diabetes Mellitus is a disease with high mortality rate that affects more than 60% population. The mindset of this task is to analyze various machine learning techniques for binary classification concerning with illness i.e. to diagnose whether a subject is suffering from disease or not. There are in total fifteen classifiers considered and out of them five major techniques namely: ANN, SVM, KNN, Naive Bayes and Ensemble are used. For achieving the desired goals the tools that were employed namely matrix laboratory (MATLAB) and WEKA 3.6.13. In Ensemble method the predictive potentials of various individual classifiers are fused together. Using Ensemble method, it increases the performance by combining the classifying ability of individual classifiers and the chances of misclassifying a particular instance are reduced significantly, this provides a greater accuracy to the overall classification process. It is the enhancing technique that does the majority voting and gives us the percolated results. The medical database analysed in this study includes a rich database of about 400 people from across a wide geographical

region and ten physiological attributes. Furthermore, this diagnostic tool is examined by verifying denary cross attestation; on top of that the outcome has been confronted along the truly existing real interpretation about the cases. A GUI based diagnostic tool founded upon ensemble classifier is developed in such a manner it would be able to predict whether a patient is enduring against the disease or not when it is fed with all the 10 attributes from user through a user friendly GUI (Graphical User Interface).The development of this diagnostic tool is done using MATLAB 2013a. Out of 10 parameters that the user needs to enter as input in GUI based diagnostic tool five are numeric and the rest are nominal values. The diagnostic tool in execution is demonstrated below in Fig. 3. The main objective of this manuscript is to propose an intelligent framework that will act as a useful aid for doctors for correct and timely biopsy can be done at early stage. The result indicated that ensemble technique assured an accuracy of 98.60% that clubs the predictive performance of multiple AI based algorithms and are superior in comparison with all other individual counterparts. The algorithms with better exactness than others are followed by Artificial neural network (ANN), Naïve Bayes, Support Vector Machine (SVM), K-Nearest Neighbor (K-NN).

EXISTING SYSTEM

Machine learning and natural language processing use big data-based models for pattern recognition, explanation, and prediction. NLP has gained much interest in recent years, mostly in the field of text analytics, Classification is one of the major task in text mining and can be performed using different algorithms

3. Since the latest data published by Johns Hopkins gives the metadata of these images. The data consists of clinical reports in the form of text in this paper, we are classifying that text into four different categories of diseases such that it can help in detecting coronavirus from earlier clinical symptoms. We used supervised machine learning techniques for classifying the text into four different categories COVID, SARS, ARDS and Both (COVID, ARDS). We are also using ensemble learning techniques for classification

PROPOSED SYSTEM

proposed a machine learning model that can predict a person affected with COVID-19 and has the possibility to develop acute respiratory distress syndrome (ARDS). The proposed model resulted in 80% of accuracy. The samples of 53 patients were used for training their model and are restricted to two Chinese hospitals. ML can be used to diagnose COVID-19 which needs a lot of research effort but is not yet widely operational. Since less work is being done on diagnosis and predicting using text, we used machine learning and ensemble learning models to classify the clinical reports into four categories of viruses.
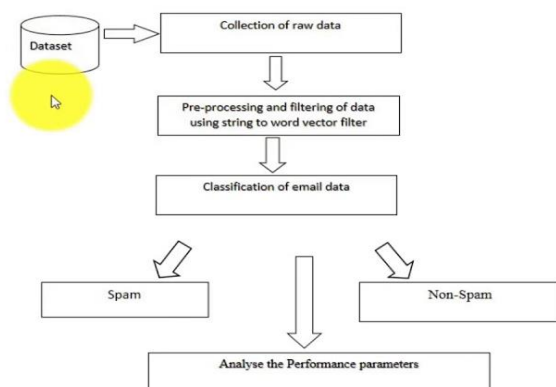
ARCHITECTURE DIAGRAM
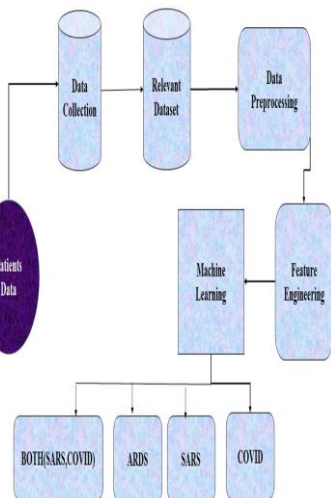
DATAFLOW DIAGRAM
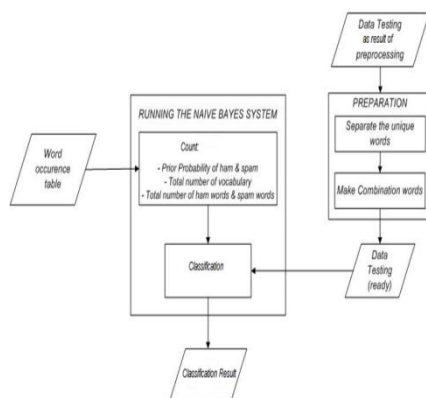
FLOW CHART

IMPLEMENTATION
INTRODUCTION

This dataset contains more than 30 columns but we are extracting two column values such as 'clinical_notes' and 'finding'. clinical_notes column contains medical text data and this text data is preprocess using NLTK library to remove stop words, special symbols and then apply lemmatizer to remove 'ing, tion etc.' from text. After preprocess text we will apply TF-IDF to extract top 40 features from dataset. Below is the dataset screen shots and this dataset saved inside 'dataset' folder.

'Upload Covid-19 Dataset' button and then upload dataset

selecting and uploading 'dataset.csv' file and then click on 'Open' button to load dataset

we extract all text data from dataset and now in above screen text in first sentence we have 'on' stop words and many number of numerical values and to remove those stop words and to clean data

preprocess all stop words removed out and in above 'on' stop word removed out

above graph showing count/finding of each label and now close above graph and then click on 'Feature Engineering' button to apply TF-IDF on above text data and to get below features

screen all text data converted to above TF-IDF features and now click on 'Run Logistic Regression, Naive Bayes, SVM & Decision Tree' to run all traditional algorithms on features data and to calculate accuracy

displaying accuracy, precision, recall and FScore for each algorithm and now click on 'Run Bagging, Adaboost, Random Forest & Gradient Boosting' button to calculate accuracy of classical algorithms

showing classical algorithms accuracy and other metrics values and now click on 'Comparative Analysis Graph' button to get below graph

we can see accuracy, precision, recall and fscore for each algorithm in group bar chart and in above graph x-axis represents algorithm name y-axis represents values
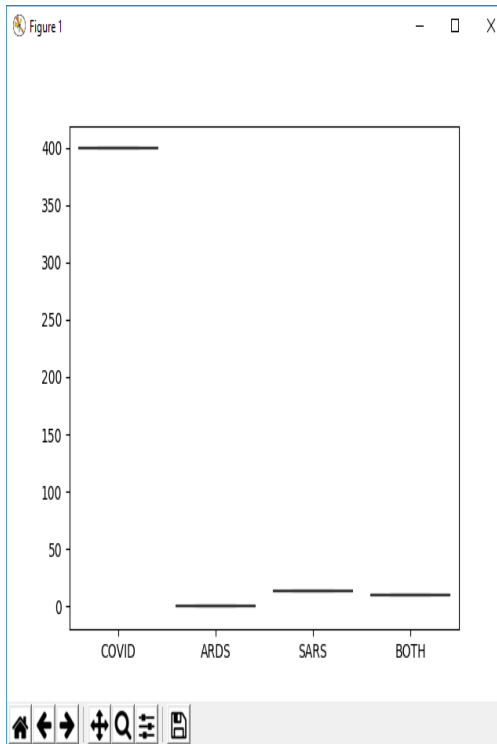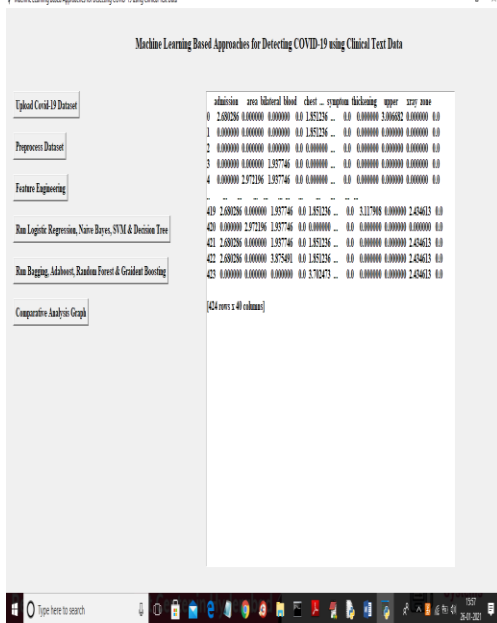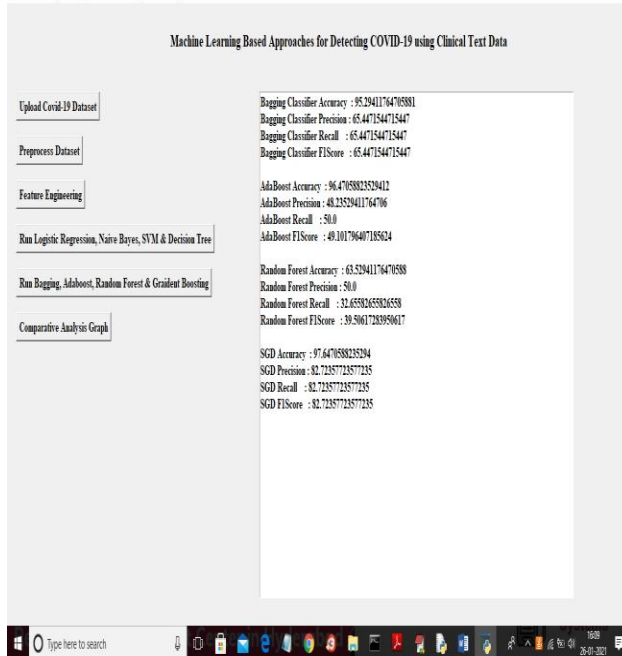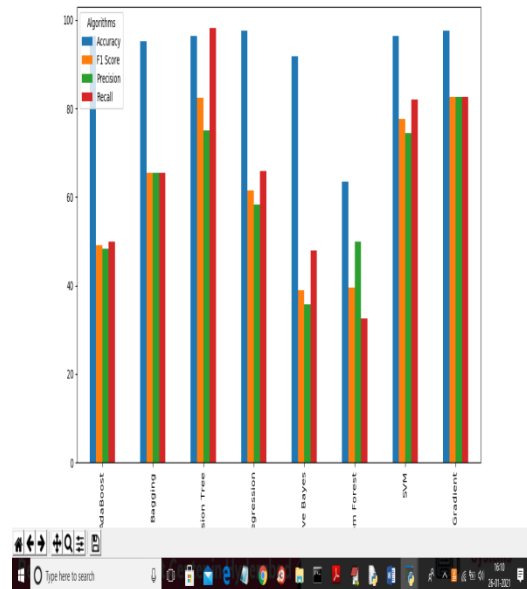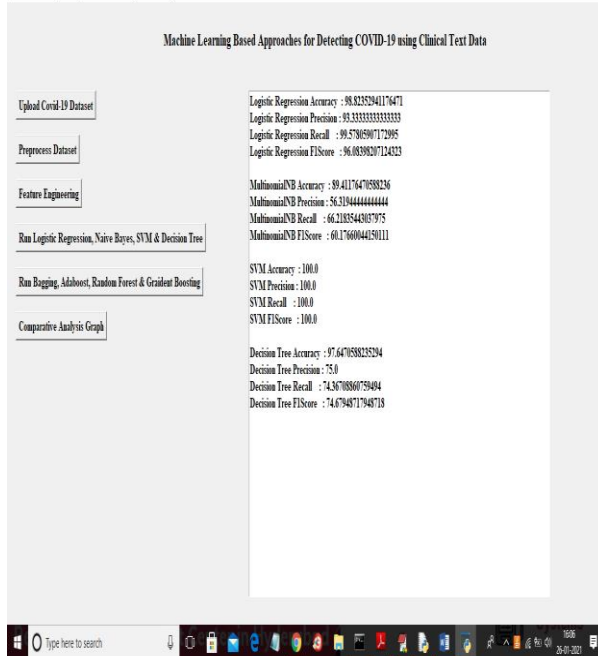
SCREEN SHOTS
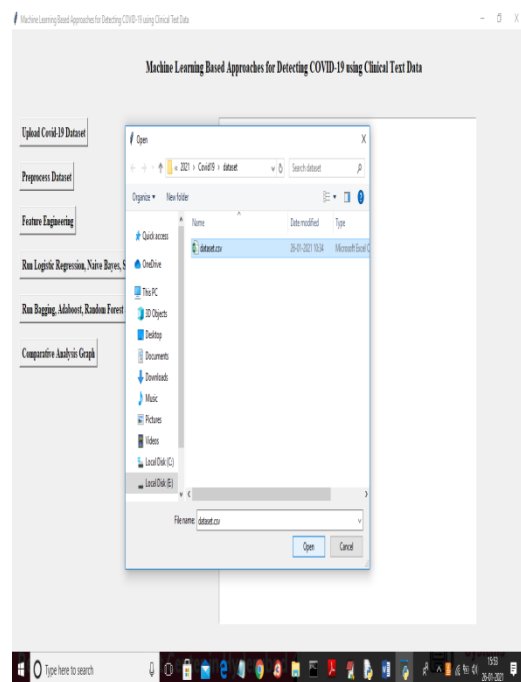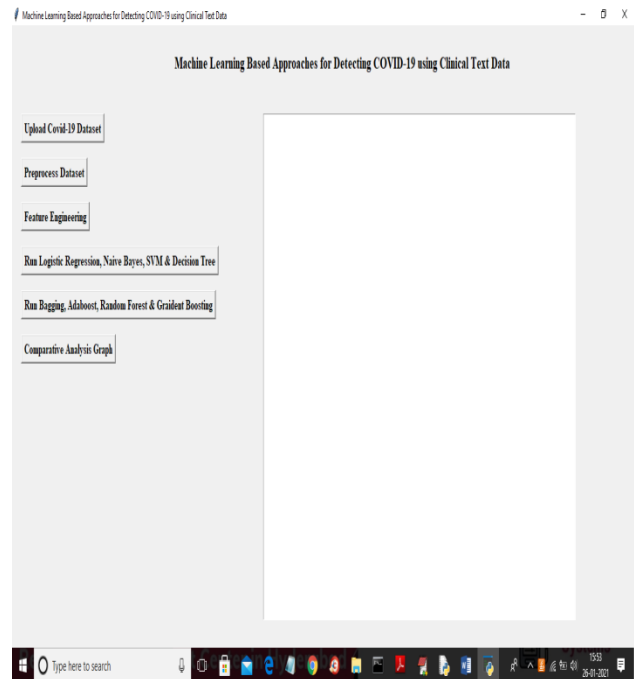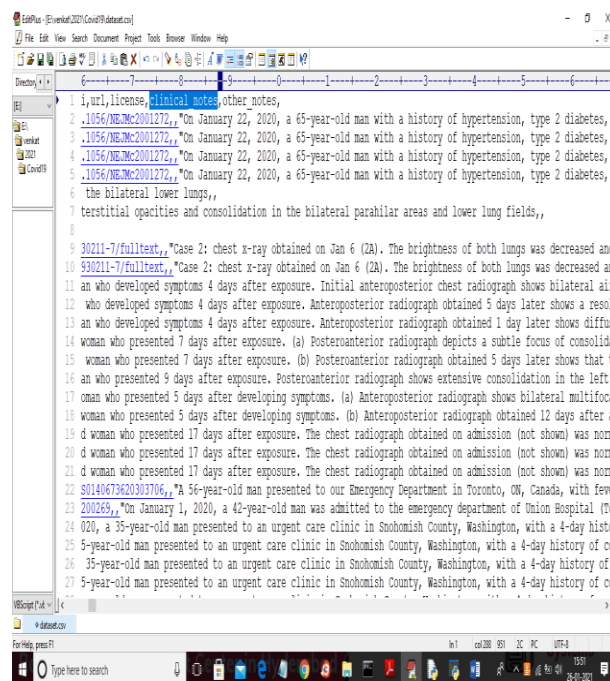
Machine Learning Based Approaches for Detecting COVID-19 using Clinical Text Data

Machine Learning Based Approaches for Detecting COVID-19 using Clinical Text Data

Upload Covid-19 Dataset

Preprocess Dataset

Feature Engineering

Run Logistic Regression, Naive Bayes, SVM & Decision Tree

Run Bagging, Adaboost, Random Forest & Gradient Boosting

Comparative Analysis Graph

Logistic Regression Accuracy : 98.8235294117647
Logistic Regression Precision : 93.3333333333333
Logistic Regression Recall : 99.5780590717299
Logistic Regression F1Score : 96.0839820712432

MultinomialNB Accuracy : 89.4117647058823
MultinomialNB Precision : 56.3194444444444
MultinomialNB Recall : 66.2183544303797
MultinomialNB F1Score : 60.1760044150011

SVM Accuracy : 100.0
SVM Precision : 100.0
SVM Recall : 100.0
SVM F1Score : 100.0

Decision Tree Accuracy : 97.6470588235294
Decision Tree Precision : 75.0
Decision Tree Recall : 74.3670886075949
Decision Tree F1Score : 74.6794871794871

Machine Learning Based Approaches for Detecting COVID-19 using Clinical Text Data

Upload Covid-19 Dataset

Preprocess Dataset

Feature Engineering

Run Logistic Regression, Naive Bayes, SVM & Decision Tree

Run Bagging, Adaboost, Random Forest & Gradient Boosting

Comparative Analysis Graph

Bagging Classifier Accuracy : 95.2941176470588
Bagging Classifier Precision : 65.4471544715447
Bagging Classifier Recall : 65.4471544715447
Bagging Classifier F1Score : 65.4471544715447

AdaBoost Accuracy : 96.4705882352941
AdaBoost Precision : 48.2352941176470
AdaBoost Recall : 50.0
AdaBoost F1Score : 49.1017964071856

Random Forest Accuracy : 63.5294117647058
Random Forest Precision : 50.0
Random Forest Recall : 32.6558265582655
Random Forest F1Score : 39.5061728395061

SGD Accuracy : 97.6470588235294
SGD Precision : 82.7285772857723
SGD Recall : 82.7285772857723
SGD F1Score : 82.7285772857723

FUTURE SCOPE AND CONCLUSION

COVID-19 has shocked the world due to its non-availability of vaccine or drug. Various researchers are working for conquering this deadly virus. We used 212 clinical reports which are labelled in four classes namely COVID, SARS, ARDS and both (COVID, ARDS). Various features like TF/IDF, bag of words are being extracted from these clinical reports. The machine learning algorithms are used for classifying clinical reports into four different classes. After performing classification, it was revealed that logistic regression and multinomial Naïve Bayesian classifier gives excellent results by having 94% precision, 96% recall, 95% f1 score and accuracy 96.2%. Various other machine learning algorithms that showed better results were random forest, stochastic gradient boosting, decision trees and boosting. The efficiency of models can be improved by increasing the amount of data. Also, the

disease can be classified on the gender-based such that we can get information about whether male are affected more or females. More feature engineering is needed for better results and deep learning approach can be used in future.

REFERENCES

1. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ (2020) A new coronavirus associated with human respiratory disease in china. Nature 44(59):265–269

2. Medscape Medical News, The WHO declares public health emergency for novel coronavirus (2020) https://www.medscape.com/viewarticle/924596

3. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X, Zhang L (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 395(10223):507–513

4. World health organization: https://www.who.int/new-room/g-adetail/ q-a-corronaviruses#:/text=symptoms. Accessed 10 Apr 2020

5. Wikipedia coronavirus Pandemic data: https://en.m.wikipedia. org/wiki/Template:2019%E2%80%9320_coronavirus_pan demic_data. Accessed 10 Apr 2020

6. Khanday, A.M.U.D., Amin, A., Manzoor, I., & Bashir, R., ''Face Recognition Techniques: A Critical Review'' 2018

7. Kumar A, Dabas V, Hooda P (2018) Text classification algorithms for mining unstructured data: a SWOT analysis. Int J Inf Technol. https://doi.org/10.1007/s41870-017-0072-1

8. Verma P, Khanday AMUD, Rabani ST, Mir MH, Jamwal S (2019) Twitter Sentiment Analysis on Indian Government Project using R. Int J Recent Tech Eng. https://doi.org/10.35940/ijrte. C6612.098319

9. Chakraborti S, Choudhary A, Singh A et al (2018) A machine learning based method to detect epilepsy. Int J Inf Technol 10:257–263. https://doi.org/10.1007/s41870-018-0088-1

10. Sarwar A, Ali M, Manhas J et al (2018) Diagnosis of diabetes type-II using hybrid machine learning based ensemble model. Int J Inf Technol. https://doi.org/10.1007/s41870-018-0270-5

11. Bullock J, Luccioni A, Pham KH, Lam CSN, Luengo-Oroz M (2020) Mapping the landscape of artificial intelligence applications against COVID-19. https://arxiv.org/abs/2003.11336v1

12. Wang L, Wong A (2020) COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 Cases from chest radiography images. https://arxiv.org/abs/2003.09871

13. Yan L, Zhang H-T, Xiao Y, Wang M, Sun C, Liang J, Li S, Zhang M, Guo Y, Xiao Y, Tang X, Cao H, Tan X, Huang N, Amd A, Luo BJ, Cao Z, Xu H, Yuan Y (2020) Prediction of criticality in patients with severe covid-19 Infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. medRxiv. https://doi.org/10.1101/2020.02.27. 20028027

14. Jiang X, Coffee M, Bari A, Wang J, Jiang X, Huang J, Shi J, Dai J, Cai J, Zhang T, Wu Z, He G, Huang Y (2020) Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. Compu Mater Contin 63(1):537–551

15. Description of Logistic Regression Algorithm. https://machine learningmastery.com/logistic-regression-for-machine-learning/. Accessed 15 May 2019

16. Description of Multinomial Naı̈ve Bayes Algorithm https://www. 3pillarglobal.com/insights/document-classification-using-multi nomial-naive-bayes-classifier. Accessed 15 May 2019

17. Khanday AMUD, Khan QR, Rabani ST. SVMBPI: support vector machine based propaganda identification. SN Appl. Sci. (accepted)

18. Description of Decision Tree Algorithm: https://dataspirant.com/ 2017/01/30/how_decision_tree_algorithm_wo rks/. Accessed 10 July 2019

19. Description of Boosting Algorithm: https://towardsdatascience. com/boosting. Accessed 10 July 2019

20. Description of Adaboost Algorithm: https://towardsdatascience. com/boosting-algorithm-adaboost-b673719ee60c. Accessed 10 July 2019

21. Katuwal R, Suganthan PN (2018) Enhancing Multi-Class Classification of Random Forest using Random Vector Functional Neural Network and Oblique Decision Surfaces, Arxiv:1802.01240v1

22. Friedman JH (2002) Stochastic gradient boosting. Comput. Stat. Data Anal. 38(4):367–378. https://doi.org/10.1016/S0167-9473(01)00065-2 Int. j. inf. tecnol. (September 2020) 12(3):731–739 739 123