Advances and Applications in Mathematical Sciences
Volume 20, Issue 12, 2021, Pages 3297-3307
© 2021 Mili Publications, India

A SWIFT EXPOSURE OF SARS-COV-2 BY MFSM MODEL

BALA BRAHMESWARA KADARU, Y. ADITYA and SIVA CHINTAIAH NARNI

Gudlavalleru Engineering College
Gudlavalleru, India
E-mail: balukadaru2@gmail.com
adityalu@gmail.com
shiva.gecmca@gmail.com

Abstract

The most epidemic of respiratory disorder is COVID-19 caused by new type of named as corona virus SARS-Cov2 is a rigorous global concern. Due to the lack of medical assistance for this situation effective treatment in the medical area the strategy of containment was the immediate action to be taken to get reduce the contagion by applying isolating the patients who are suffering with this virus. On the other hand isolation cannot be completely resolve and practically difficult to put in practice for a long time. To take faster decisions on healing with isolation will give better features to conclude for suspected infection cases. These predicted cases can be the best predictors for positive analysis. To do the analysis on patient characteristics, symptoms, diagnosis and outcomes. By using machine learning supervised algorithms a model is implemented to recognize the features of COVID-19 disease with best accuracies. Many supervised machine learning methods were been studied for the dataset considered in our analysis. The MFSM algorithm performed with the highest accuracy (>94%) to predict COVID-19 status for all age groups. Statistical analysis of for the collected symptoms is fever (41.1%), cough (30.3%), lung infection (13.1%) and runny nose (8.43%). Out of which 54.4% of patients did not have any symptoms.

I. Introduction

The severe acute respiratory syndrome corona virus SARS-CoV2 corona virus [1] (World Health Organization) is been rapidly growing which is leading to a respiratory disease COVID-19. First time the virus was been recognized in 1960's as 229E and OC43. Later different individual virus types

2010 Mathematics Subject Classification: 68.
Keywords: SARS-Cov-2, COVID-19, Corona virus, Supervised Learning, respiratory disorder.
Received November 20, 2020; Accepted December 20, 2020

were identified from this group of virus and found to be from the zoonotic infections is common in different bat families. The infections that occurred in the respiratory system was been seemed to be harmless. With the growth of frequency of severity and toxic disorders in the respiratory system lead to initial corona virus. In past two decades more number of cases were observed in severe acute respiratory symptoms (SARS) and the middle east respiratory symptoms (MERS). In the year 2002, Fushun a place in China SARS-CoV found with mortality rates of 9.6% and Saudi Arabia in 2012 MERS-CoV with mortality rates of 36% from diagnosed patients found to be first rapidly spread of infection places, which gave raise to panic situation in worldwide and brought the situation of containment zones. The corona virus infections identified are noteworthy threat to human health as it is portable to lead serious respiratory infections in humans, especially if one-to-one infection occurs quickly.

The corona virus [2] rate of growth and speed of spread causing COVID-19 this lead to fast up development of vaccine and therapeutic. However, the first observations within weeks of COVID-19 virus, which was isolated and characterized. The structure of SARS-CoV2 protein targets is a 3C-like protease. More efforts are kept for Clinical tests for drugs and virtual screening for possible targets using protein structure data [3]. Infected individuals are given priority to isolate and give treatment based on level of severity. To identify the patients easily with clinical symptoms. Based on the studies conducted in the past, within the state of Andhra Pradesh it was identified as COVID-19 by the medical tests and death rate conditions. It was also assumed to be the reason for fastest spread of disease routes through respiratory transmission.

The fatal growth analyzed for the COVID-19 was 1.4% in China, in February 2020. On the other hand, it is quite challenging to estimate the vastly accurate from different country to country. If suppose we observe Italy during March 2020, the fatality rate stood to be 7.2% [4].This is how it reflects social differences in between the nations, where 23% of the Italians whose age is more than 65. When compared to china the Italians rate of infections remain high when stratified by age i.e. over 70 years of age. This leads to the critical need to improve transmission and detection procedures for classifying severe attack rate of the given population which changes

among the nations. Here, we will use Supervised Learning methods as these are well suited for improving the identification rate of infected patients who are having COVID-19 positive symptoms. By considering COVID-19 data from various major populated places in Andhra Pradesh was validated for predictions of pandemic case. The following are the major techniques of machine learning:

- The Dataset will be constructed based on the data which was collected from hospitals those who are admitted as in-patients and extracted the features which are useful for diagnosis.
- Five different supervised machine learning techniques were used to identify the COVID-19 patients based on their symptoms.
- We develop a model which is useful to predict COVID-19 attacked persons among supposed and confirmed patients considering few factors such as their age, past travelling history.

II. Related Works

In 1965, corona virus was initially identified in human with cold as the major complaint. B814 name was given to that virus which could not explore in society. In a related study of Procknow and Hamre, the authors recognized new type of virus and it was named as 229E. They recognized and isolated the virus based on the samples which were collected from the medical students who are suffering with cold. In a further research done by McIntosh and Dees the atmosphere sensitive agents of different strains were inaccessible from human respiratory region. Since they were developed in the organ culture, hence named as OC.

At that time Almeida JD, Tyrrell DA calculated organ features affected by B814 detailed with electron microscopy carried with size 80 to 150. Later, another type of viruses were identified, named as CORONA.

The various viruses carried from China between 2002-03, recognized as SARS corona virus and spread among various countries.

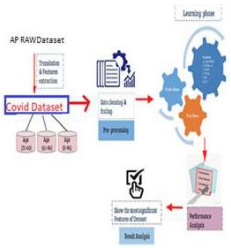
III. COVID-19 Disease Data

Data Collection

The raw hospital data is collected from various Health camp conducted from various centers (COVID-19-tracker, 2020). Individual's records are made to be available at hospitals and clinics for further diagnosis and treatment. The repositories gathered within Andhra Pradesh from various districts were around 10,000 samples. The datasets collected was again verified with World Health Organization for further processing. Due to the exploit of corona virus, the epidemiology information was used for further analysis. The repositories collected were structured for analysis whose features are elementary information like age, work, gender, work, residential history; past information like time, place, transportation and event performed. These elementary information was used for detecting the COVID-19 disease.

Table 1- Feature descriptions

Feature	Type	Description
Gender	String	Almost same ratio of male and female patients
Age	Integer	The age range is 0–96 years
Fever	Boolean	Develops symptoms with a high body temperature of 38 °C or more
Cough	Boolean	Develops symptoms with a dry cough or cough with sputum
Pneumonia	Boolean	Develops symptom of pneumonia and admitted to hospital
Lung Infection	Boolean	Radiographic or CT scan indicates chest imaging changes as lung infection
Runny Nose	Boolean	Develops the symptom of runny nose
Muscle Soreness	Boolean	Develops symptoms of limb or muscle soreness



The datasets of Andhra Pradesh for COVID-19 includes information about all the patients those who were confirmed as positive as well as suspected to be positive from the patients who have symptoms and had contact with COVID-19 patients. After diagnosis, it is found that all the patients are not confirmed as COVID-19 patients. The COVID-19 will be confirmed based on the CDC test report by the doctors in the root dataset. Finally, dataset is prepared with features which are described in Table1. From the 10,000 samples, the conformed COVID-19 cases were 2,000 and 8,000 were suspected cases. The symptoms of all patients were varied like diarrhea muscle soreness as such.

IV. Ensemble Decision Tree Model

The models of Feature selection are divided into two groups: Feature Subset Selection and Feature Ranking. On the training dataset different statistical measures are evaluated and subset of features is selected by using

Feature selection. The model of the classification is as shown in Figure 1.

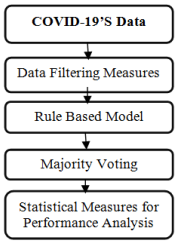


Figure 1: Traditional ensemble disease prediction

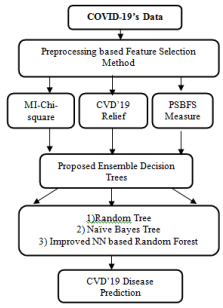


Figure 2: Model of COVID'19 disease Prediction

Proposed Multi-layered Ensemble Decision Tree classification Model: the proposed classification model is implemented on the test data to detect the impact of the virus by using ensemble classifiers. The data ambiguity and class inequity are the developing areas of research in the prediction of better accuracy rate in virus identification. The below diagram describes the selection procedure of feature subset.

Step1. Data pre-processing on Training CVD'19 data.

Input. CVD'19 data

$$CVD^1, CVD^2, \ldots, CVD^n.$$

For each feature: A (i) in the

$$CVD^1, CVD^2, \ldots, CVD^n$$

do

For each tuple $I(j)$ in the $A(i)$ do.

If (is Numerical $(I(j))$ and $I(i) = \text{empty})$ then

$$I(j) = \frac{\sum_{j=1/i=1}^{n} (|\ \text{Max}\ \{A(I(j))\} - \text{Mode}_{A(I(j))})\ |)}{(\text{Max}_{A(I(j))} - \text{Min}_{A(I(j))})} * \text{Sealing\_factor} \quad (1)$$

end if

if (is Categorical A_i) and $A_i(I)$ = empty)

then

$P(j)$ = Prior Prob ($A(i)$, class (m));

$$I(j) = \frac{\sum_{j=1/i\neq j}^{n} (|\ \text{Max}\ \{p(j)\} - \text{Min}\ \{P(j)\}\ |)}{\text{Sealing} - \text{factor}}. \tag{2}$$

Sealing − factor $\in (0, 1)$.

Here, m^{th} class of the missing value is used to find the prior probability in place of missing value

end if

End for

PSBFSM Technique [5]:

The PSBFSM Technique can be implemented in three modules. They are:

(1) PSM Technique (2) Chi-Square Technique (3) Correlation Technique

1. PSM Technique:

CS measure is used to find out the class distributions mapping with the matching instances of disease. This method origin is KL-divergence. In equation (3), we derived the Formulae for PSM as

$$PSM = \sqrt[3]{\left(\sum_{i=1}^{|D_i\|D_i|}\sum_{j=1}^{|D_i|}(\sqrt[3]{D_i/|\ D_i\ |} - \sqrt[3]{D_j/|\ D_j\ |})^2\right)} * \sum_{i=1}^{|D_i\|D_i|}\sum_{j=1}^{|D_i|}\text{Prob}\,(D_i) * \log\,(\text{Prob}\,(D_i/\text{Prob}\,(D_i)) \tag{3}$$

where D_i is infected disease class, D_j is normal class.

2. Chi-Square Technique:

This test is used to categorical feature in data set, which is used identify expected virus affected patients in the observed unaffected patients.

3. Correlation Technique:

Correlation is an analysis that measures the strength of association between the positive and negative relationship of the COVID diseases

$$\text{MIChi} - \text{square}\,(D_i,\, D_j) = \text{Max}\,\{P(D_i) * \log\,(D_i/D_j),\, \sum\sum (D_i - D_o)^3/D_o\}$$

$$\text{Yates Correction}\,(\text{MIChi} - \text{square}) = \sum\sum (|\, D_i,\, D_o\,| - 0.5)^2/D_o. \qquad (4)$$

CVD Relief (COVID'19 Distance Relief Feature Selection):

Let CVD+ and CVD- are +ve , – ve instances

$$WCVD = (0,\, 0,\, \ldots,\, 0);\; DCVD = \text{Norm}\,(D,\, 0,\, 1).$$

For each attribute F in CVD do

For each tuple in CVD do

Randomly select tuple T€CVD

Randomly Consider +ve tuple closest to CV_{D^+} as

$$P^+ € CVD^+.$$

Randomly Consider -ve tuple closest to CV_{D^+} as

$$P^- € CVD^-.$$

If $(T € CVD^+)$.

Then

$$H^+ = \rho^+;\; \text{hitrate}\; H^+ = \rho^-;\; \text{missrate}$$

Else

$$H^+ = \rho^-;\; \text{hitrate}\; H^+ = \rho^+;\; \text{missrate}.$$

Done

Using distance measure hit rate and miss rate are been updated with weights

$$W_i = \text{Max}\{\text{Normalize}(H^+, H^-)\} * (W_i - \sqrt{(R - H^+) + (R - H^-)^2})$$

$$\text{Threshold}(\lambda_i) = (1/m)\sum_{i=1}^{n} W.$$

If ($\lambda_i \geq u$ threshold).

Then

Select feature F.

ADD to CVDF List (F)

Else

Remove feature F.

Done

MFSED-Tree construction [6].

Input: Selected Feature Set CVDFList [];

Output: Disease patterns

Procedure:

Read Feature Dataset CVDFList

For each Feature CVDF[i] in CVDFList

Do

For each instance $I(A_i)$ in A_i do

Do

Divide the data instances of CVDF (D_i) into 'k' independent sets.

Select classifier $C_{i/i=1, \ldots, m}$ {Naïve Bayes, Random Tree, Proposed Decision Tree}

Load CVD Dataset

(a) Replacement of Training Data with Test Data

(b) Use Equation (6) to construct the decision tree

$$RFFSM = -\frac{\sqrt[3]{PD\text{Relief}(D_i, D_j) * | D | * PSM(D_i, D_j)}}{\text{Yates Corr (MI – Chisquare)}} \quad (6)$$

(c) Apply sorting on tree to separate affected and unaffected people

(d) Apply majority voting for node selection using ensemble classifiers.

End while

Calculate error rate, accuracy, TP, FP and F-Measure

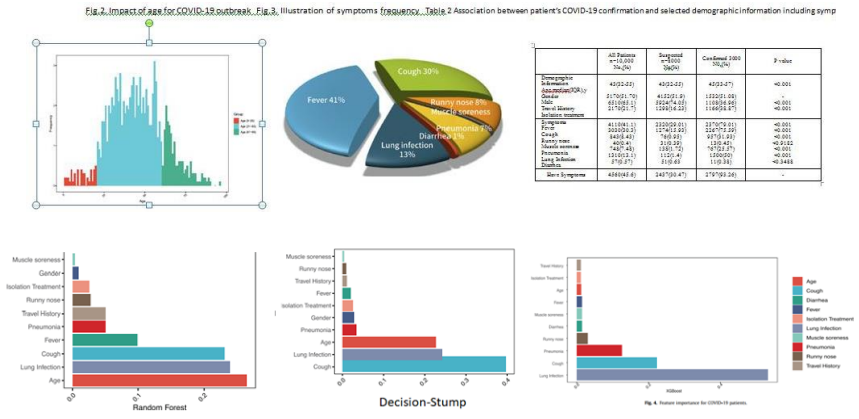
Done

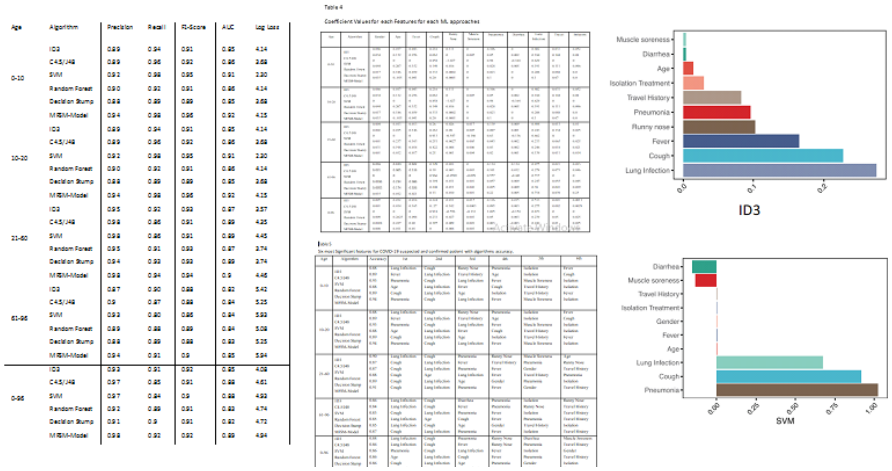
Done

By using this model, we will optimize the construction of the decision tree.

V. Experimental Results

The Experimental analysis was also been observed by using SPSS Modeler 16.0, weka and Orange. Age attribute of each member has been considered as an average of 43 years, IQR was 32 to 55 years for 5,170 males (51.7%). The observations of patient COVID-19 confirmation and relevant symptoms statistical information are shown in Table 3.





Discussion and Conclusion

The fast growth of pandemic COVID-19 is the current issue throughout the world health so, the current action for each researcher is to identify the COVID-19 symptoms in each of the member as early as possible so that we can stop spreading it from one to one. For this process to put in practice machine learning approach is used to help this find the symptoms in pandemic situation. Various machine learning approaches were studied to predict COVID-19.The significant symptoms found were runny nose, fever, lung infection, and cough, pneumonia, past travel, gender isolation, age, muscle soreness and diarrhea. In the machine learning approaches studied with the above attributes brought up with good accuracies to find the stage of COVID-19. With the MFSM algorithms accuracy for the age range 10–20 years, was 94% accuracy, where as other approaches performed with more than 85% accuracy. The age 21 to 60 years was with same case with accuracy of 91% of MFSM and others.

The machine learning algorithms were been developed on various symptoms of patients with COVID-19 infection in a different dataset from Andhra Pradesh and applied various classifiers to analyze the results in its performance. These results will assist the health centers to predict COVID-19 infection in advance by the features used in the analysis. Even though some of the classifiers were not so accurate in predicting the pandemic situation. Nevertheless, for the analysis of COVID'19 dataset considered for analysis

may not be enough to satisfy completely So, as the future scope its advised to work out with more feature and larger records for still better analysis in this pandemic case of COVID-19 and which may also increase in the accuracies for predicting.

References

[1] R. Agarwal, The 5 Classification Evaluation metrics every Data Scientist must know. https://towardsdatascience.com/the-5-classification-evaluation-metrics-you-must-know-aa97784ff226 Accessed 18 April 2020. https://www.aljazeera.com/news/2020/01/countries-confirmed-cases-coronavirus-200125070959786.html Accessed 18 April 2020.

[2] S. Tian, N. Hu, J. Lou, K. Chen, X. Kang and Z. Xiang, et al., Characteristics of COVID-19 infection in Beijing. Journal of Infection 80 (2020), 401-406. https://doi.org/10.1016/j.jinf.2020.02.018.

[3] A. Zhavoronkov, V. Aladinskiy, A. Zhebrak, B. Zagribelnyy, V. Terentiev and D. S. Bezrukovet al. Potential 2019-nCoV 3C-like Protease Inhibitors Designed Using Generative Deep Learning Approaches, BioRxiv. https://doi.org/10.26434/chemrxiv.11829102.v1 (2020).

[4] G. Onder, G. Rezza and S. Brusaferro, Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. Jama. https://doi.org/ 10.1001/jama.2020.4683. (2020).

[5] Bala Brahmeswara Kadaru and B. Raja Srinivasa Reddy, A novel ensemble decision tree classifier using hybrid feature selection measures for Parkinson's disease prediction, Int. J. Data Science 3(4) (2018), 289-307.

[6] Bala Brahmeswara Kadaru and B. Raja Srinivasa Reddy, An improved parallel PSO-FS based N-SVM technique for medical disease prediction, J. Adv. Res. Dynam. Control Sys. 10 (2018), 723-737.