ISSN 2515-8260 Volume 09, Issue 07, 2022

Computer Aided Covid-19 Mortality Scope Prediction by Supervised Learning

[*1]Monelli Ayyavaraiah

[1]Research Scholar, Dayananda Sagar University-Bangalore, Assistant Professor, Department of Computer Science and Engineering, KKR & KSR Institute of Technology & Sciences (KITS), Vinjanampadu, Vatticherukuru Mandal, Guntur-522017.

[*1]Email: ayyavarayya.monelli@gmail.com

[2]Dr. Bondu Venkateswarlu

[2]Associate Professor, Dayananda Sagar University-Bangalore.

Abstract: The infection of the covid-19 fears the world population. As prevention strategies have not been developed, existing clinical approaches are only applicable to treat covid-19-positive individuals. Identifying the severity of the patient's illness is crucial for reducing the covid-19-related mortality rate. It is the pathology reports that are used as the foundation for determining the severity of the disease by the clinical specialists. However, a clinician's skill in making a diagnosis has a significant impact on how correct that diagnosis turns out to be. This manuscript described a supervised learning technique for performing computer-assisted covid-19 mortality scope using the pathology reports of the target patient. The experimental examination of the value of the suggested approach for anticipating mortality scope with few false alarms.

Keywords: *Machine Learning, COVID-19, Computed Tomography, Feature Optimization, Tailed-Test, Supervised Learning, Medical Imaging.*

1 Introduction

There are numerous epidemic diseases, which attacked human beings over the past several years. To fight over these pandemics, the WHO (world health organization) is cooperating with several national authorities as well as clinicians. As, the first case of COVID-19 disease has been confirmed and occurred in Wuhan district, china in the year December 2019, it spread all over the world and finally, on January 30th 2020, it has been declared by WHO that this pandemic disease is an international concern as stated in [1].

The work in [2], [3] presents that, COVID-19 has been considered as an infectious disease, which is caused by new coronavirus and identified in WUHAN city, china. The term SARS-CoV-2 (severe acute respiratory syndrome coronavirus-2) is a novel type of virus, which has not been recognized in people earlier. Moreover, the virus occurs mainly through respiratory issues, droplets through sneezing, coughing or while people gets interacted over each other. When these droplets are inhaled or they might land on surfaces and other people might touch with their hand and people get contaminated to this virus if their hand comes in contact with their nose, eyes or mouth.

The coronavirus, which has become an epidemic disease might live on diversified surfaces like plastic and stainless steel for few days and in case of copper and cardboard the virus might live for some hours. Nevertheless, the amount of possible virus may fall off after some time and

would not always exist to cause infection. Moreover, in instance of humans, the virus symptoms can be observed between 1-14 days starting from infection day. Later, its starts spreading with high speed, providing no time for preparing against novel detected notorious and infectious virus that obliged the WHO in declaring COVID-19 to be more pandemic as stated in [4] because of rapid transmission among humans. Furthermore, several people have already got infected in the world and numerous lives have been lost.

Various trails in the clinical department have been going on for measuring the effectiveness of covid-19, however, there are no results available till date. Since, it is a novel virus, there is no availability of vaccine also. Even though, several pharmaceutical and research companies have started to work and research on vaccine, it might take months or year before the vaccine is available for humans to use [5]. Each and every one should be aware of this virus and take necessary precautions in overcoming it.

In section 1, the introduction related to COVID-19 has been discussed along disease spreading all over the world. In section 2, several approaches carried by various researchers has been discussed. Section 3 explores the details of the data, methods, and materials used in the proposed statistical assessment method that predicts the mortality scope of the patient tested positive to covid-19. Section 4 exhibits the simulation or experimental outcomes followed by conclusion in section 5.

2 Related work

The accurate outbreak estimation methods have to be accessed for attaining insights and spread the causes of disease causing. The other legislative and government bodies depend on insights from estimation methods for recommending the novel policies and for measuring the enforced rules effectiveness as stated in [6].

The work [7] presents that, the contemporary global pandemic disease COVID-19 has been envisioned complex and non-linear in nature. Moreover, the epidemics has variances with other contemporary epidemics that brings to raise a question to understand the standard methods capability for delivering accurate outcomes as stated in [8]. Also, several unknown as well as known variables included in the spread, the intricacy of extensive population behavior in several geo-political domains and variances in containment schemes had augmented method uncertainty intensely as stated in [9]. Accordingly, the standard epidemiological methods face novel challenges for delivering more consistent outcomes. For overcoming this challenge, various novel methods have evolved that introduce various assumptions towards modelling as stated in [10-12].

The machine learning has been utilized for enhancing the screening procedure and diagnosis of recognized patient through radio imaging scheme similar to clinical data of blood sample and CT (computed tomography).

The expert of healthcare uses radiology images such as CT scans and X-ray as routine devices for enhancing the conventional screening and diagnosis. Inappropriately, such devices performance is moderate at the time of maximal SARS-CoV-2 pandemic outburst. In respect to this, the work [13] exhibit possible ML devices by recommending a novel method, which comes with valid as well as rapid SARS-CoV2 diagnosis model.

The current studies design an ancillary device for enhancing the COVID-19 diagnosis accuracy with novel Automatic COVID-19 identification model by depending on algorithm called deep learning as stated in [14]. The proposed method utilizes raw images of chest x-ray of 127 patients who are infected. With a superior accuracy of performance, the 98.08% of binary class as well as 87.02% of multiclass has been attained. The multi-class envisioned the expert system applicability for assisting the radiology in verifying the screening procedure accurately and rapidly. Moreover, several researchers have identified four significant combinations of medical features in laboratory, clinical features as well as demographic information by utilizing percentage of CD3, GHS, overall protein and age of patient employing SVM as significant feature classification method as stated in [15]. Also, the novel method is robust and effective in estimating patients in severe or critical conditions, and simulation outcomes exhibited that integration of 4 features results an AUROC of 0.9757 and 0.9996 in testing and training datasets respectively.

After assessing 253 clinical samples of blood from Wuhan, several researchers have identified 11 significant relevant indices that support as discrimination device of covid-19 over the healthcare proficient towards quick diagnosis as stated in [16]. The contributions exhibited that 11 relevant indices have been extracted after engaging random forest (RF) algorithm with 95.95% of accuracy and specificity of 96.97% respectively.

The AI and ML applications have been used in predicting and estimating the existing pandemic situation. The novel approach, which estimated and predicted in 1-6 days of overall 10 patients in states of brazil is by using stacking-ensemble through a SVM regression algorithm on accumulative positive cases of COVID-19 from Brazilian data. Hence, increases the short-term prediction procedure for alerting the healthcare expert and government for handling pandemic as stated in [17]. The contemporary studies recommended a novel approach by utilizing supervised recursive multi-layered classifier known as XGBoost on mammographic and clinical parameter datasets. Later on, implementing the method, the researchers identified three prominent features from 75 clinical features and samples of blood test result as 90% of accuracy in estimating and measured the COVID-19 patient as general, moderate and severe as stated in [18].

The decision rule has been implemented for forecasting quickly and estimating the infected people at maximum risk, the patients who are declared as infected should be considered for intensive caring, and mainly reduce the shortness rate. The forecasting approach based on Canadian by using time-series has been developed by engaging deep-learning algorithm over an extensive short-term network of memory. The work [19] presents that several researchers have been identified a significant parameter aimed for estimating the course with a prediction of ending point of current epidemic SARS-CoV2 in Canada as well as in the entire universe. The recommended approach prediction ending point of SARS-CoV2 pandemic in Canada would be approximately june 2020

The work [20] presents that, depending on gathered data from the University of John Hopkins, the estimation is likely precise as a novel infected case have been reduced rapidly. The realistic forecasting method has been projected by integrating goodness of forecasting model based on wavelet and further, the time series approach based on auto-regressive integrated moving average as depicted in [21].

The objective of the contemporary contributions is the prediction of the covid-19 scope from the clinical reports of the target patient. Though the computer aided clinical practices are

ISSN 2515-8260 Volume 09, Issue 07, 2022

certainly essential to improve the sensitivity and specificity of the COVID-19 prediction, the severity (mortality scope) of the diseased patient prediction is crucial to treat the patient with customized course of medical recommendations. Concerning to the objective of mortality scope prediction, this manuscript proposed a statistical analysis scale.

The contemporary contribution Gradient Boosting Survival Model (GBSM) [22] has explored the performance of diversified supervised learning methods to predict the mortality scope of diseased individual, who tested positive in covid-19 test. However, only attributes age and sex has recommended as optimal to train the classifier, and the classification process more centric to predict the time to discharge. The other contribution Individual-Level Fatality Prediction Model (ILFPM) [23] that motivated from the discharge time prediction by Gradient Boosting Survival Model (GBSM) [22] has endeavored to identify the scope of artificial intelligence methods to predict the individual fatality scope of the patients tested positive to the covid-19. However, these contributions are using the demographic features, which includes any chronic disease is positive or negative. The other demographic features related to health condition haven't been used in these contemporary contributions.

Contrast to these methods, proposed method has used the diversified features related to clinical diagnosis outcomes of the patients to predict the fatality/ mortality scope of an individual, who tested positive in covid-19 test.

3 Methods and materials

This section explores the details of the data, methods, and materials used in the proposed statistical assessment method that predicts the mortality scope of the patient tested positive to covid-19. The block diagram representation of C19MP is represented in Figure 1.

3.1 The data

The demographic features that includes the features like age, gender, weight, and clinical reports related to pathological reports related to blood tests. The details considered from clinical reports are subset of the pathology reports relevant to diabetes type-I and type-II. Each record contains 31attributes, which includes the mortality scope either positive or negative.

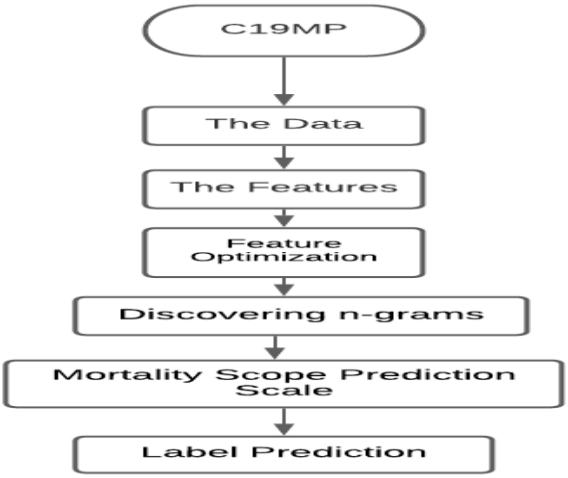

Figure 1: The block diagram representation of C19MP

3.2 The features

The training dataset tD bipartite as two sets tD_+, tD_- contains the records labeled positive and negative respectively. Let the set aL contains all attributes (except mortality scope) representing the values projected in each record of the both sets tD_+, tD_-. Further discovers all possible subsets of size 1 to size $|aL|$ of the set aL. These subsets are referred further as n-gram feature labels and the values projected in the records for each n-gram feature label referred as n-gram features. The frequency of each n-gram feature in datasets tD_+, tD_- denotes their confidence towards the labels positive and negative respectively. However, discovering the n-grams is complex and proportionate to feature attribute's count. Hence, reducing the attribute's count optimizes the volume of n-grams to be used in learning phase.

3.3 Feature Optimization

The values projected to an attribute having significant diversity between their confidence towards the labels positive and negative are derived as optimal attributes. The t-test [24], which is a distribution diversity assessment measure has used to discover the diversity between the values of an attribute in records labeled as positive and the values of respective attribute in records labeled as negative. The mathematical model of feature optimization explored in following description.

$\forall_{i=1}^{|aL|}\{a_i \exists a_i \in aL\}$ Begin //for each attribute a_i of the attribute list aL

Let the vectors af_i^+, af_i^- denote the values of the attribute a_i in positive and negative label sets tD_+, tD_-

$\left\langle af_i^+ \right\rangle = \frac{1}{|af_i^+|}\left(\sum_{j=1}^{|af_i^+|}\{e_j \exists e_j \in af_i^+\}\right)$ // mean of the values listed in the vector af_i^+

ISSN 2515-8260 Volume 09, Issue 07, 2022

$\left\langle af_i^- \right\rangle = \frac{1}{|af_i^-|}\left(\sum_{j=1}^{|af_i^-|}\left\{e_j \exists e_j \in af_i^-\right\}\right)$ // mean of the values listed in the vector af_i^-

$\sigma\left(af_i^+\right) = \frac{1}{\left|af_i^+\right|}\left(\sum_{j=1}^{|af_i^+|}\left\{\sqrt{\left(\left\langle af_i^+ \right\rangle - e_i\right)^2}\right\}\right)$ // deviation of the values listed in vector af_i^+

$\sigma\left(af_i^-\right) = \frac{1}{\left|af_i^-\right|}\left(\sum_{j=1}^{|af_i^-|}\left\{\sqrt{\left(\left\langle af_i^- \right\rangle - e_i\right)^2}\right\}\right)$ // deviation of the values listed in vector af_i^-

$ts = \frac{\left(\left\langle af_i^+ \right\rangle - \left\langle af_i^- \right\rangle\right)}{\sqrt{\sigma\left(af_i^+\right) + \sigma\left(af_i^-\right)}}$ The t-score discovered from the vectors af_i^+, af_i^- representing values of the attribute a_i towards positive and negative labels respectively.

Further, probability value [25] pv of the t-score ts shall discover from the t-table [26]

$if\ (pv < p\tau)$ If the probability value pv is lesser than the probability threshold $p\tau$

Indicates both vectors as divergent, hence the attribute a_i is having optimal values to consider in learning phase

$else$ $aL \setminus a_i$ if the probability value pv is not less than the probability threshold $p\tau$, discard attribute a_i from the list of attributes aL, which is since the attribute a_i is not reflecting diversity between the it's values in both sets tD_+, tD_- of positive and negative label.

End // of the $\underset{i=1}{\overset{|aL|}{\forall}}\left\{a_i \exists a_i \in aL\right\}$

Further phase, discovers the optimal n-gram features for both the labels as explored in following section.

3.4 Discovering n-grams

The resultant set aL is used as input to discover all possible unique subsets, which denotes as n-gram features in further descriptions. The mathematical model to discover n-grams of dynamic sizes explored in following description.

$nGr \leftarrow aL$	The set aL is default set of 1-grams. Hence, the attributes listed in set aL shall move to set nG of n-grams
$tnG \leftarrow nGr$	The set tnG is the clone of the set nGr
$while\left(\left\|tnG\right\| > 0\right)$ Begin	While the set tnG is not empty
$\underset{i=1}{\overset{\|tnG\|}{\forall}}\left\{ng_i \exists ng_i \in tnG\right\}$ Begin	For each n-gram ng_i of the set tnG
$\underset{j=1}{\overset{\|tnG\|}{\forall}}\left\{ng_j \exists ng_j \in tnG \wedge i \neq j\right\}$ Begin	For each n-gram ng_j of the set tnG that does not equal to the n-gram ng_i
$ng \leftarrow \left\{ng_i \cup ng_j\right\}$	New n-gram ng, which is the union of two n-grams ng_i, ng_j that

		denotes
$if\,(ng \notin nGr)$	$nGr \leftarrow ng$	If n-gram ng does not exist in the set nGr of n-grams, add n-gram ng to the set nGr
End		// of the loop $\underset{j=1}{\overset{\|tnG\|}{\forall}}\{ng_j \exists ng_j \in tnG \wedge i \neq j\}$
End		// of the loop $\underset{i=1}{\overset{\|tnG\|}{\forall}}\{ng_i \exists ng_i \in tnG\}$
$if\,(\|nGr\| > \|tnG\|)$ Begin		If the size $\|nGr\|$ of the set nGr is greater than the size $\|tnG\|$
$tnG \setminus tnG$		Empty the set tnG
$tnG \leftarrow nGr$		Add all the n-grams of the set nGr to the empty set tnG
End		Of the condition $if\,(\|nGr\| > \|tnG\|)$
$else\,if\,(\|nGr\| \equiv \|tnG\|)$	$tnG \setminus tnG$	If both the sets nGr, tnG having same size, empty the set tnG
End		// of the loop $while(\|tnG\| > 0)$

The resultant set nGr contains all possible subsets of attributes (feature labels) of the set aL. Further, discovers the n-gram feature values and their confidence towards both sets tD_+, tD_- as follows

$\underset{i=1}{\overset{\|nGr\|}{\forall}}\{ng_i \exists ng_i \in nGr\}$ Begin	// for each n-gram ng_i feature of the set nGr
$fv(ng_i)$	// is an empty set containing unique n-grams of feature values of the n-gram feature ng_i both positive and negative labels
$\underset{j=1}{\overset{\|tD_+\|}{\forall}}\{r_j \exists r_j \in tD_+\}$ Begin	// for each record r_j of the set tD_+
$fv(ng_i) \leftarrow \{v(ng_i) \exists v(ng_i) \subseteq r_j \wedge v(ng_i) \notin fv(ng_i)\}$	values $v(ng_i)$ that is subset of record r_j of positive label and does not exists in set $fv(ng_i)$ of the n-gram feature ng_i
End	// of the loop $\underset{j=1}{\overset{\|tD_+\|}{\forall}}\{r_j \exists r_j \in tD_+\}$
$\underset{j=1}{\overset{\|tD_-\|}{\forall}}\{r_j \exists r_j \in tD_-\}$ Begin	// for each record r_j of the set tD_- of records labeled as negative
$fv(ng_i) \leftarrow \{v(ng_i) \exists v(ng_i) \subseteq r_j \wedge v(ng_i) \notin fv(ng_i)\}$	values $v(ng_i)$ that is subset of record r_j of negative label and does not exists in set $fv(ng_i)$ of the n-gram feature ng_i
End	// of the loop $\underset{j=1}{\overset{\|tD_-\|}{\forall}}\{r_j \exists r_j \in tD_-\}$
End	// of the loop $\underset{i=1}{\overset{\|nGr\|}{\forall}}\{ng_i \exists ng_i \in nGr\}$

#Finding the positive and negative confidence of each n-gram feature values#

$\underset{i=1}{\overset{\|nGr\|}{\forall}}\{ng_i \exists ng_i \in nGr\}$ Begin	// for each n-gram ng_i feature of the set nGr
$\underset{j=1}{\overset{\|fv(ng_i)\|}{\forall}}\{v_j \exists v_j \in fv(ng_i)\}$	Each n-gram feature value v_j of n-gram feature ng_i

$pc_+ \leftarrow \frac{1}{|tD_+|}\left(\sum_{k=1}^{|tD_+|}\{1 \exists v_j \subseteq r_k)\}\right)$ Move positive confidence of n-gram feature value v_j

$pc_- \leftarrow \frac{1}{|tD_-|}\left(\sum_{k=1}^{|tD_-|}\{1 \exists v_j \subseteq r_k)\}\right)$ Move negative confidence of n-gram feature value v_j

End // of the loop $\overset{|fv(ng_i)|}{\underset{j=1}{\forall}}\{v_j \exists v_j \in fv(ng_i)\}$

End // of the loop $\overset{|nGr|}{\underset{i=1}{\forall}}\{ng_i \exists ng_i \in nGr\}$

3.5 Mortality Scope Prediction Scale

The coefficients using to scale the mortality scope shall discover as projected in following mathematical model. Coefficients are the mean and respective root-mean-square-deviation (RMSD) of the positive and negative confidence of the diversified n-gram feature values.

$\langle pc_+ \rangle = \frac{1}{|pc_+|}\left(\sum_{i=1}^{|pc_+|}\{c_i \exists c_i \in pc_+\}\right)$ // The mean of predictive confidence of the n-gram feature values of the positive label.

$\langle pc_- \rangle = \frac{1}{|pc_-|}\left(\sum_{i=1}^{|pc_-|}\{c_i \exists c_i \in pc_-\}\right)$ // The mean of predictive confidence of the n-gram feature values of the negative label.

$pc_+^{\sigma} = \frac{1}{|pc_+|}\left(\sum_{j=1}^{|pc_+|}\left\{\sqrt{(\langle pc_+ \rangle - c_j)^2} \exists c_j \in pc_+\right\}\right)$ // The deviation *RMSD* of the predictive confidence of the n-grams of the positive label.

$pc_-^{\sigma} = \frac{1}{|pc_-|}\left(\sum_{j=1}^{|pc_-|}\left\{\sqrt{(\langle pc_- \rangle - c_j)^2} \exists c_j \in pc_-\right\}\right)$ // The deviation *RMSD* of the predictive confidence of the n-grams of the negative label.

$lb_+ = \langle pc_+ \rangle - pc_+^{\sigma}$ //The lower-bound of the **positive prediction scale** is the difference between mean and respective deviation of the positive confidence of n-gram feature values

$ub_+ = \langle pc_+ \rangle + pc_+^{\sigma}$ // The cumulative of mean and deviation denotes the upper-bound of the prediction scale of the **positive label**.

$lb_- = \langle pc_- \rangle - pc_-^{\sigma}$ //The lower-bound of the **negative prediction scale** is the difference between mean and respective deviation of the positive confidence of n-gram feature values

$ub_- = \langle pc_- \rangle + pc_-^{\sigma}$ // The cumulative of mean and deviation denotes the upper-bound of the prediction scale of the **negative label**.

Return (ss, ssl, ssu) // Returns n-gram coefficient, upper as well

as lower Heuristics regression boundaries.

3.6 Label Prediction

Let the input record r has been given to predict the mortality scope is positive or negative. Collect all the n-gram feature values of all optimal n-gram features from the given record r. Find the mean of the positive and negative confidence $\langle c_+ \rangle, \langle c_- \rangle$ of all n-gram feature values of the given input record r towards training corpus.

Further, predict the mortality scope of the record as follows,

- $if\left(\langle c_+ \rangle \geq ub_+\right)$// If the mean confidence $\langle c_+ \rangle$ of the n-gram feature values discovered from the given input record r is greater than or equals to positive upper-bound ub_+, the mortality scope against the given record r is positive
- $if\left(\langle c_+ \rangle \geq \langle pc_+ \rangle \&\& \langle c_- \rangle < \langle pc_- \rangle\right)$ // the condition mean confidence $\langle c_+ \rangle$ is greater than the mean of the positive confidence $\langle pc_+ \rangle$ and mean confidence $\langle c_- \rangle$ is lesser than the negative confidence $\langle pc_- \rangle$ also recommends to label the given record as positive (prone to mortality).
- $\langle c_+ \rangle \geq lb_+ \&\& \langle c_- \rangle < lb_-$ // The condition that denotes mean $\langle c_+ \rangle$ of the confidence discovered for positive label of given record r is greater than lower-bound lb_+ of the positive label prediction scale and mean $\langle c_- \rangle$ of the negative confidence of the n-gram feature values of given record r is lesser than the lower-bound lb_- of the negative label prediction scale

Further, down the conditions contradicting to above stated conditions, the given input record shall be labeled as negative (not prone to mortality scope).

4 Experimental Study

This experimental investigation evaluated suggested as well as existing methods using demographic and pathological characteristics of covid19-positive individuals. Accuracy, specificity, sensitivity, precision, Mathew's correlation coefficient (MCC), and f-measure were used to assess the quality of these proposed and existing models in a cross validation procedure. The outcomes observed from this metrics exhibits the performance advantage of the projected C19MP model when compared to other existing or counterpart methods. The works [27], [28] presented Gradient Boosting Survival Model (GBSM) and Individual-Level Fatality Prediction Model (ILFPM) respectively to discover the mortality scope of the covid-19 patients, which have considered to compare with and scale the performance of the proposed model.

4.1 The dataset

The dataset considered for experimental study has records of 5000 equally shared between two labels positive and negative. Each of the record listed in dataset is having demographic and pathology features of the anonymized patient. The total number of features including label are 31. The optimal feature selection process discovered 17 features as optimal among the 30 features (excluding the label).

ISSN 2515-8260 Volume 09, Issue 07, 2022

4.2 Results Discussion

Table 1: Mean values of the cross-validation metrics

	C19MP	GBSM	ILFPM
Positive predictive value	0.9203±0.004796	0.89961±0.004257	0.88624±0.006422
True Negative Rate	0.89812±0.006397	0.87287±0.005889	0.85659±0.008803
True Positive Rate	0.90675±0.008177	0.87811±0.005502	0.86118±0.004322
Prediction Accuracy	0.903±0.005864	0.87585±0.004175	0.85919±0.005471
False alarming	0.097±0.0058642	0.12415±0.0041754	0.14081±0.005471
F-measure	0.90907±0.00561	0.88605±0.005087	0.87115±0.007662
MCC	0.80326±0.011731	0.74865±0.008393	0.71515±0.011289

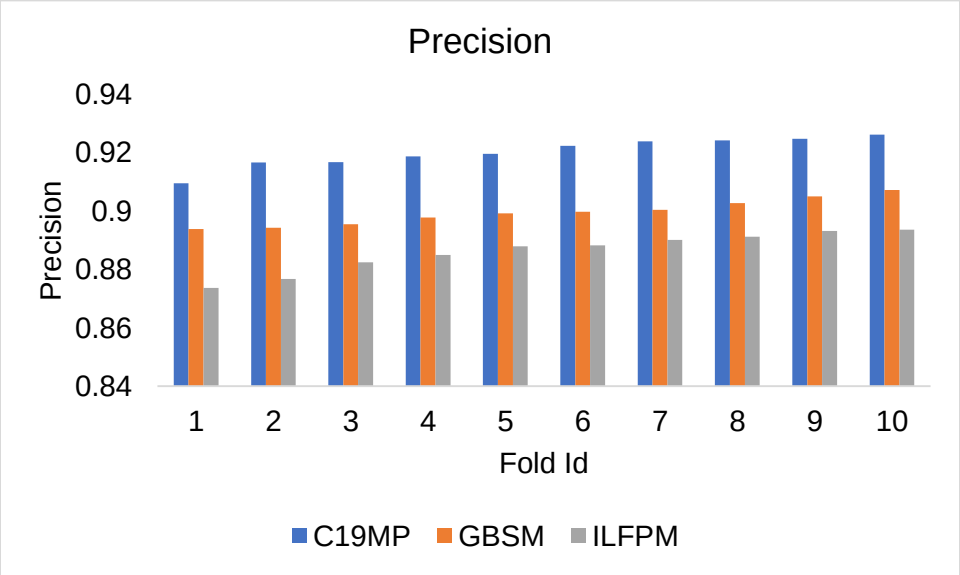

Figure 2: Positive predictive value (precision) obtained from 10-fold cross validation over proposed C19MP and contemporary models GBSM and ILFPM

The proposed model C19MP has been compared with other counterpart GBSM and ILFPM as reflected in figure 2. The graph in figure 2, Table 1 has been plotted among precision over 10-folds of cross validation observed from the proposed and contemporary models. The average precision for the GBSM, ILFPM, and proposed C19MP are 0.89961±0.004257, 0.88624±0.006422, and 0.9203±0.004796 respectively. The performance of the contribution C19MP is therefore more notable and favorable toward positive predictive value when compared to contemporaries GBSM as well as ILFPM models, according to the data of the precision stated above.

ISSN 2515-8260 Volume 09, Issue 07, 2022

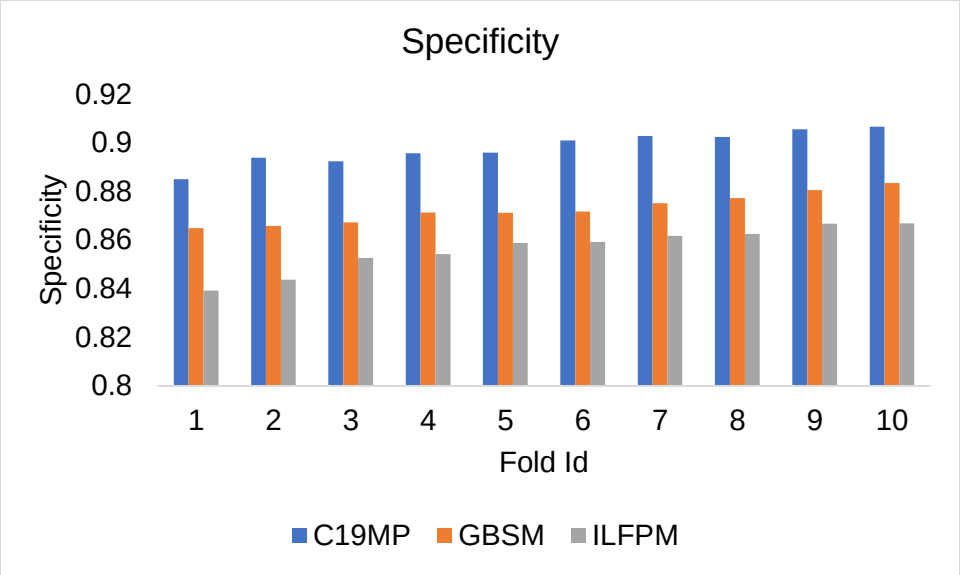

Figure 3: Specificity from 10-fold cross validation of C19MP, GBSM, and ILFPM

This metric specificity denotes how optimal the classifier is in detecting the negative cases. In this empirical study, this metric has been measured to exhibit the performance of C19MP and counter models. In figure 3, comparison among specificity over 10-fold cross validation of C19MP, ILFPM and GBSM has been represented in the form of graph. The average specificity of the projected C19MP model, and counterpart GBSM and ILFPM model are 0.89812±0.006397, 0.87287±0.005889 and 0.85659±0.008803 respectively. Thus, it has been envisioned from the above discussion that, the true negative rate of the method C19MP is superior that compared to counterpart methods.

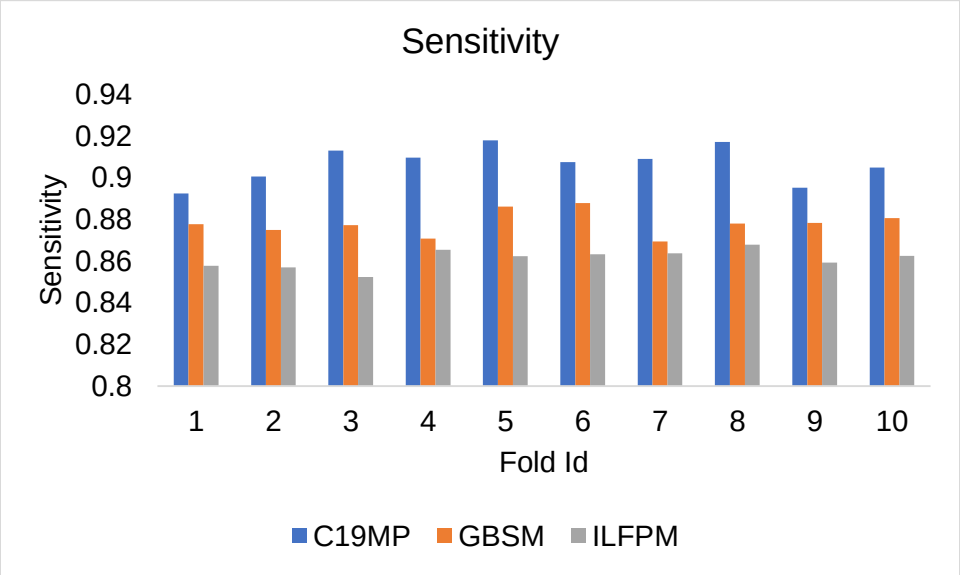

Figure 4: Sensitivity from 10-fold cross validation of C19MP, GBSM, and ILFPM

ISSN 2515-8260 Volume 09, Issue 07, 2022

The sensitivity that reflects true positive rate of the contribution C19MP has been compared with the sensitivity observed from counterpart models GBSM and ILFPM in figure 4. The graph has been plotted among sensitivity observed from 10-folds cross validation for the proposed and contemporary models. The average sensitivity for the GBSM, ILFPM and C19MP are 0.87811±0.005502, 0.86118±0.004322, and 0.90675±0.008177 is respectively. These statistics exhibiting that, the performance of the contribution C19MP is more prominent and advantageous than the counterparts GBSM and ILFPM models.

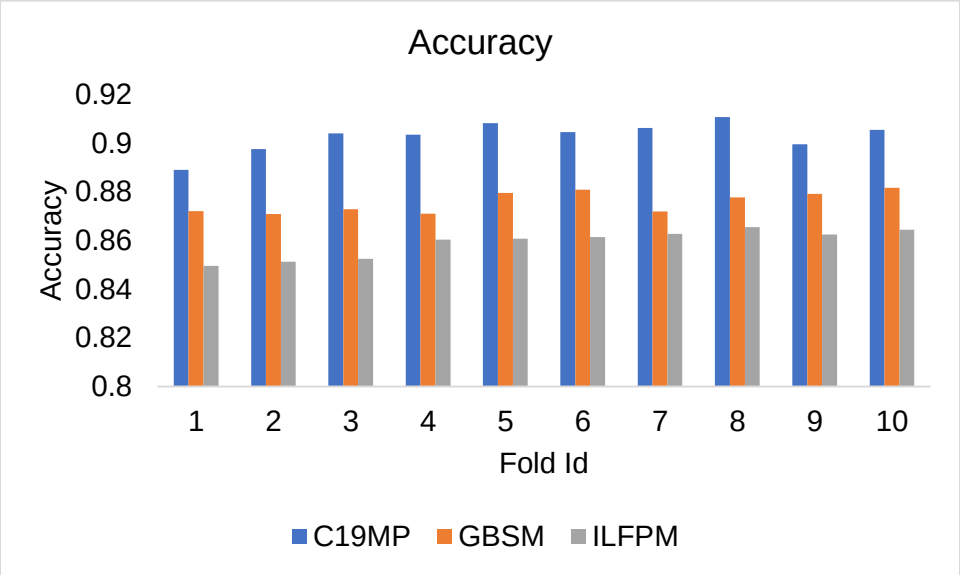

Figure 5: Accuracy from 10-fold cross validation of C19MP, GBSM, and ILFPM

The metric accuracy denotes the accurate prediction rate of both labels. In figure 5, comparison among accuracy obtained from 10-fold cross validation of C19MP, ILFPM and GBSM has been represented in the form of graph. The average accuracy of the contribution C19MP model, and counterparts GBSM and ILFPM models are 0.903±0.005864, 0.87585±0.004175 and 0.85919±0.005471 respectively. Accordingly, it may be inferred from the foregoing description that the C19MP performs more accurately than competing approaches.

Figure 6 shows a line graph of F-measures generated from cross validation between the suggested as well as existing models, reflecting a comparison of C19MP with other comparable models like GBSM as well as ILFPM. The average F-measure of the GBSM is 0.88605±0.005087 and for ILFPM is 0.87115±0.007662. The average F-measure of the contribution C19MP is 0.90907±0.00561. From the F-measure data shown above, it follows that the suggested C19MP is more salient and favorable than its counterparts, the GBSM as well as the ILFPM models.

ISSN 2515-8260 Volume 09, Issue 07, 2022

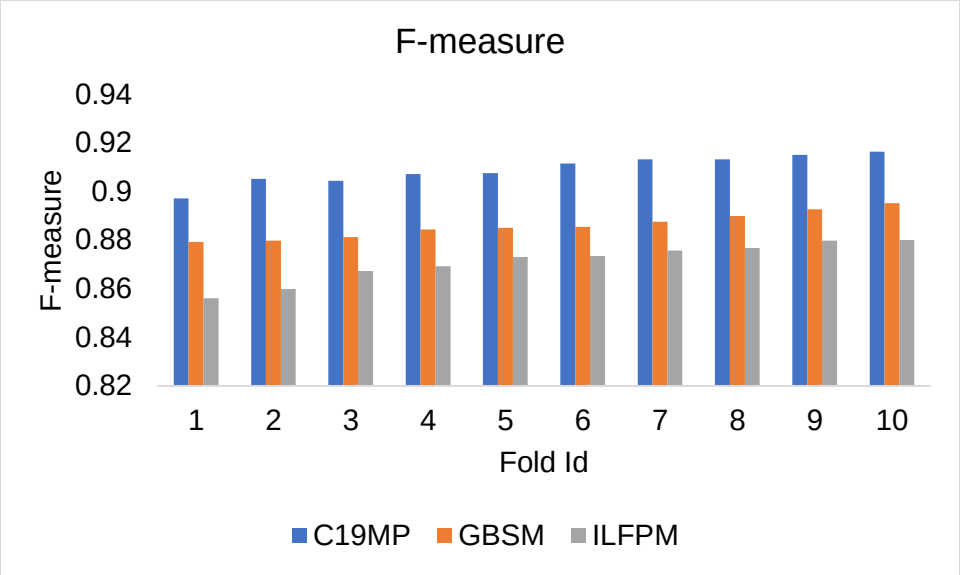

Figure 6: F-Measure from 10-fold cross validation of C19MP, GBSM, and ILFPM

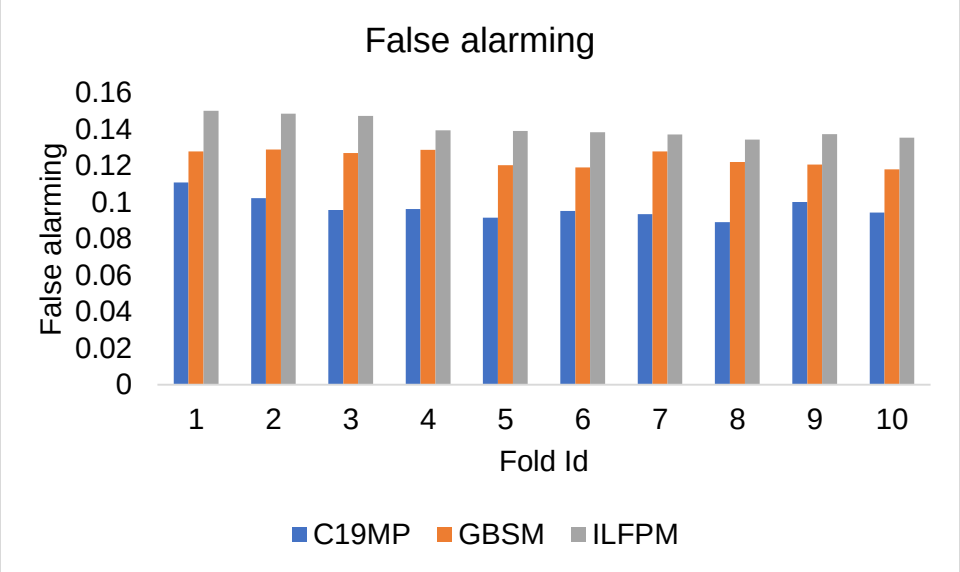

Figure 7: False Alarming from 10-fold cross validation of C19MP, GBSM, and ILFPM

The forecast failure rate is indicated by the measure false alarming. Figure 7 shows the cross validation of the suggested C19MP, as well as existing ILFPM, and GBSM's false alarms. The average false alarming of the contribution C19MP, and counterparts GBSM and ILFPM are 0.097±0.0058642, 0.12415±0.0041754 and 0.14081±0.005471 respectively. From the foregoing discussion, it has been concluded that the proposed approach of this article performs better than competing techniques in terms of minimum false alarms.

ISSN 2515-8260 Volume 09, Issue 07, 2022

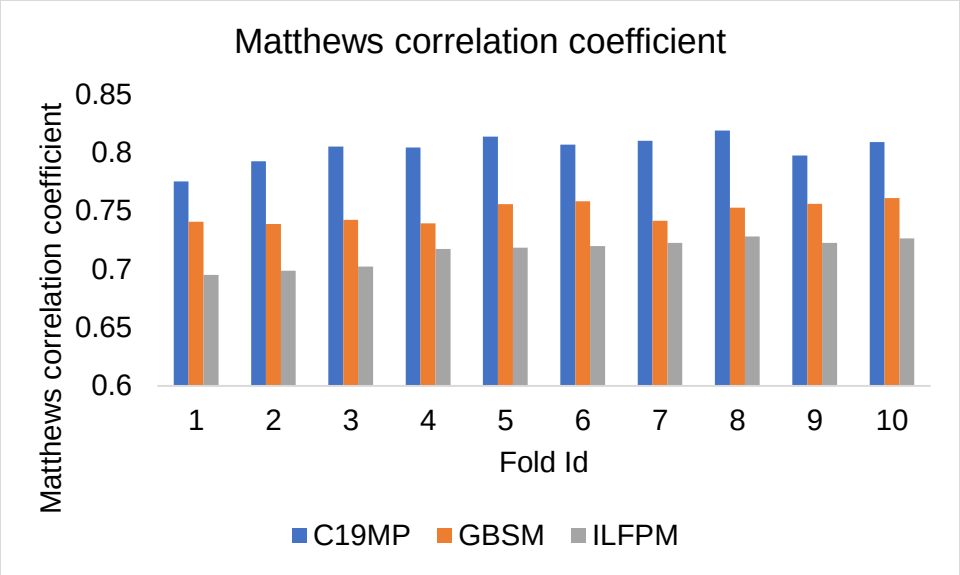

Figure 8: MCC from 10-fold cross validation of C19MP, GBSM, and ILFPM

The GBSM and ILFPM 10-fold cross validation averages of MCC are 0.74865±0.008393 and 0.71515±0.011289 respectively. This C19MP's MCC on average is 0.80326±0.011731. Therefore, it can be inferred from the data stated above that the performance of suggested C19MP is superior to that of counterpart GBSM and ILFPM models.

5 Conclusion

A novel statistical method has been proposed to predict mortality scope of a patient tested positive for covid-19. Unlike contemporary models, which considered only demographic features, the proposed model has considered the features of demographic and pathology reports to perform supervised learning. In order to reduce the process complexity of learning phase, a novel feature optimization measure has portrayed that build on distribution diversity method t-test. The significance of the contribution C19MP has exhibited by the 10-fold cross validation carried on recommended dataset. The proposed model C19MP has scaled by comparing with contemporary methods of recent past. The future research can use the cross-media features, which is the combination of demographic, image, and signal formats to improve the prediction accuracy of the mortality scope of the patient tested positive for COVID-19.

References

[1] CS, ZA, M. K. N ON, and AAJ AK. "World Health Organization declares global emergency: A review of the 2019 novel." Int J Surg 76 (2020): 71-76.

[2] Zhu, Na, et al. "A novel coronavirus from patients with pneumonia in China, 2019." New England journal of medicine (2020).

[3] Tan, Wenjie, et al. "A novel coronavirus genome identified in a cluster of pneumonia cases—Wuhan, China 2019− 2020." China CDC Weekly 2.4 (2020): 61-62.

[4] Gorbalenya, Alexander E., et al. "Severe acute respiratory syndrome-related coronavirus: The species and its viruses–a statement of the Coronavirus Study Group." (2020).

[5] Mode of transmission by country, person infectious2020b; https://www.ecdc. europa.eu/en/covid-19/questions-answers.

[6] Remuzzi, Andrea, and Giuseppe Remuzzi. "COVID-19 and Italy: what next?." The lancet 395.10231 (2020): 1225-1228.

[7] Ivanov, D. "Predicting the impacts of epidemic outbreaks on global supply chains: a simulation-based analysis of the COVID-19/SARS-CoV2 case." Transp. Res. E https://doi. org/10.1016/j. tre (2020).

[8] Koolhof, Iain S., et al. "The forecasting of dynamical Ross River virus outbreaks: Victoria, Australia." Epidemics 30 (2020): 100377.

[9] Darwish, Ali, Yasser Rahhal, and Assef Jafar. "A comparative study on predicting influenza outbreaks using different feature spaces: application of influenza-like illness data from Early Warning Alert and Response System in Syria." BMC research notes 13.1 (2020): 1-8.

[10] Rypdal, Martin, and George Sugihara. "Inter-outbreak stability reflects the size of the susceptible pool and forecasts magnitudes of seasonal epidemics." Nature communications 10.1 (2019): 1-8.

[11] Scarpino, Samuel V., and Giovanni Petri. "On the predictability of infectious disease outbreaks." Nature communications 10.1 (2019): 1-8.

[12] Zhan, Zhicheng, et al. "Real-Time Forecasting of Hand-Foot-and-Mouth Disease Outbreaks using the Integrating Compartment Model and Assimilation Filtering." Scientific reports 9.1 (2019): 1-9.

[13] Ardakani, Ali Abbasian, et al. "Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks." Computers in Biology and Medicine 121 (2020): 103795.

[14] Song, Zelai, et al. "Effects of microfluid in the veins of the deployable hindwings of the Asian ladybeetle on flight performance." Computers in biology and medicine 121 (2020): 103817.

[15] Sun, Liping, et al. "Combination of four clinical indicators predicts the severe/critical symptom of patients infected COVID-19." Journal of Clinical Virology 128 (2020): 104431.

[16] Wu, Jiangpeng, et al. "Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results." MedRxiv (2020).

[17] Ribeiro, Matheus Henrique Dal Molin, et al. "Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil." Chaos, Solitons & Fractals 135 (2020): 109853.

[18] Yan, Li, et al. "An interpretable mortality prediction model for COVID-19 patients." Nature machine intelligence 2.5 (2020): 283-288.

[19] HU: John Hopkins University, 2020. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). https://www.coronavirus.jhu.edu/map.html (accessed 09 June 2020)

[20] Chimmula, Vinay Kumar Reddy, and Lei Zhang. "Time series forecasting of COVID-19 transmission in Canada using LSTM networks." Chaos, Solitons & Fractals 135 (2020): 109864.

[21] Chakraborty, Tanujit, and Indrajit Ghosh. "Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis." Chaos, Solitons & Fractals 135 (2020): 109850.

[22] Nemati, Mohammadreza, Jamal Ansary, and Nazafarin Nemati. "Machine-learning approaches in COVID-19 survival analysis and discharge-time likelihood prediction using clinical data." Patterns 1.5 (2020): 100074.

[23] Li, Yun, et al. "Individual-level fatality prediction of COVID-19 patients using AI methods." Frontiers in Public Health 8 (2020): 566.

[24] Budak H, Taşabat SE. A modified t-score for feature selection. Anadolu Üniversitesi Bilim Ve Teknoloji Dergisi A-Uygulamalı Bilimler ve Mühendislik. 2016 Dec 20;17(5):845-52.

[25] Sahoo P, Riedel T. Mean value theorems and functional equations. World Scientific; 1998.

[26] Retrieved from http://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf. Accessed on 11/08/2017.

[27] Alfaras M, Soriano MC, Ortín S. A fast machine learning model for electrocardiogram-based heartbeat classification and arrhythmia detection. Frontiers in Physics. 2019; 7:103.

[28] Mustaqeem A, Anwar SM, Majid M. Multiclass classification of cardiac arrhythmia using improved feature selection and SVM invariants. Computational and mathematical methods in medicine. 2018;2018.