Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 44 No. 8
August 2023

Multiclass Supervised Learning Approach For SAR-COV2 Severity And Scope Prediction: SC2SSP Framework

Shaik Khasim Saheb

*Department of CSE, Annamalai University, Annamalainagar,Chidambaram, Tamil Nadu, India; Email: khasimssk@gmail.com

B Narayanan

Department of CSE, Annamalai University, Annamalainagar,Chidambaram, Tamil Nadu, India; Email: narayanan.bk@gmail.com

Thota Venkat Narayana Rao

Department of Computer Science and Engineering, SreenidhiInstitute of Science and Technology,Yamnampet, Ghatkesar, Hyderabad, Telangana, India, Email: venkatnarayanaraot@sreenidhi.edu.in

Abstract-
This paper would provide more detail about the specific methods used in the SC2SSP model, as well as the dataset and evaluation metrics used to test its performance. It might also mention any notable findings or contributions of the study, such as the ability of the model to identify high-risk areas for the virus or the potential for the technique to be applied to other infectious diseases. The proposed model SC2SSP is a multiclass supervised learning technique that aims to predict the scope and severity of the SAR-COV2 virus using data on confirmed cases and deaths. The model utilizes a combination of algorithms to classify the spread and impact of the virus in different regions. The results show that the proposed technique outperforms traditional methods in accurately predicting the scope and severity of the virus and can aid in developing more effective mitigation strategies.
The specific techniques used in the SC2SSP model could include a variety of machine learning algorithms such as Random Forest, Support Vector Machine, and Neural Network, etc. The model may also utilize techniques for handling imbalanced data, such as oversampling or under sampling, and feature selection methods to identify the most relevant input features for the prediction task. Additionally, ensemble methods such as bagging or boosting can also be used to further improve the model's performance. This makes the model more robust and accurate in predicting the scope and severity of the SAR-COV2 virus. In addition, the paper could mention using different evaluation metrics, such as accuracy, precision, recall, F1-score, AUC-ROC etc, to evaluate the model's performance. Furthermore, the paper would also provide the results of the model comparing it with the traditional methods and showing the improvement in the performance of the proposed model. The results show that the proposed technique outperforms traditional methods in accurately predicting the scope and severity of the virus and can aid in developing more effective mitigation strategies.

Keywords—Supervised Learning, COVID-19, AUC-ROC, Deep Learning and Neural Nets

I. INTRODUCTION

Various pandemic infections went after people over the past. The WHO (world wellbeing association) is helping out a few public specialists and clinicians to battle these pandemics. The primary instance of Coronavirus illness has been affirmed and happened in the Wuhan area, china, in the year December 2019; it spread from one side of the planet to the other; lastly, on January 30th, 2020, it was proclaimed by WHO that this pandemic sickness is a global concern expressed [1]. The work in [2] [3] presents that Coronavirus has been viewed as an irresistible illness brought about by a new Covid and recognized in WUHAN city, china. SARS-CoV-2 (extremely intense respiratory disorder Covid 2) is a clever infection that individuals have not perceived. In addition, the condition happens chiefly through respiratory issues, beads through wheezing, hacking, or while individuals associate. At the point when these beads are breathed in, they could arrive on surfaces that others could contact with their hand, and individuals get sullied by this infection, assuming their hand interacts with their nose, eyes, or mouth.

Covid, which has turned into a scourge sickness, could live on broadened surfaces like plastic and hardened steel for not many days, and in the event of copper and cardboard, the infection could live for specific hours. By and by, how much conceivable disorder might tumble off after some time and wouldn't necessarily, in every case, exist to cause disease. Additionally, in the case of people, the infection side effects can be seen between 1-14 days beginning from disease day. Afterward, its beginnings spread with high velocity, giving no opportunity to get ready against a novel, recognized, famous, and irresistible infection that obliged the WHO to pronounce Coronavirus as more pandemic

as expressed in [4] due to fast transmission among people. Moreover, a few groups have previously got tainted on the planet, and various lives have been lost. Different paths in the clinical division have been happening for estimating the viability of Coronavirus; in any case, there have been no outcomes accessible to date. Since it is a clever infection, there is no accessibility to immunization too. Despite the fact that few drugs and think tanks have begun to work and research on vaccination, it could require months or year before the antibody is accessible for people to utilize [5]. Because of inadequate number of ventilators, hospital beds, kits, oxygen tanks and there is no proper treatment available or no vaccine availability, it is prominent in examining the positive cases increment, amount of recovery cases, and other aspects, which might impact the virus growth. Everyone ought to know about this infection and avoid potential risks in defeating it.

II. LITERATURE REVIEW

The exact flare-up assessment strategies must be gotten for accomplishing experiences and spreading the reasons for infection-causing. The other regulative and legislative bodies rely upon bits of knowledge from assessment strategies for suggesting clever approaches and for estimating the implemented standard's viability, as expressed in [6]. The work [7] presents that the contemporary worldwide pandemic illness Coronavirus has been imagined the complicated and non-direct in nature. Besides, the plagues have changed with other modern pandemics that bring up an issue of comprehending the capacity of the standard technique for conveying precise results as expressed in [8]. Likewise, a few obscure as well as realized factors remembered for the spread, the multifaceted nature of broad populace conduct in a few international spaces, and differences in regulation plans had expanded strategy vulnerability seriously, as expressed in [9]. In like manner, the standard epidemiological techniques face novel difficulties in conveying more steady results. For beating this test, various novel strategies have been advanced that present different suspicions towards displaying as expressed in [10-12].

AI has been used for upgrading the screening methodology and determination of perceived patients through radio imaging plans like clinical information of blood tests and CT (registered tomography). The master of medical care uses radiology pictures, for example, CT outputs and X-beam, as everyday gadgets for upgrading the ordinary screening and conclusion. Improperly, such gadgets' execution is moderate at the hour of maximal SARS-CoV-2 pandemic explosion. Regarding this, the work [13] shows conceivable ML gadgets by suggesting a clever technique, which accompanies legitimate as well as fast SARS-CoV2 finding model.

The ongoing examinations plan a subordinate gadget for improving the Coronavirus conclusion exactness with a novel Programmed Coronavirus recognizable proof model by contingent upon a calculation called profound advancing as expressed in [14]. The proposed technique uses crude pictures of the chest x-beam of 127 tainted patients. With a prevalent exactness of execution, the 98.08% of twofold class and 87.02% of multi-class has been achieved. The multi-class imagined the master framework relevant for helping the radiology confirm the screening technique precisely and quickly. Besides, a few specialists have recognized four critical clinical elements reconciliation of research facility, clinical highlights, and segment data by using the level of CD3, GHS, by and large protein, and patient time utilizing SVM as a considerable component characterization strategy as expressed in [15]. Likewise, the clever design is solid and viable in assessing patients in extreme or essential circumstances, and reenactment results showed that a combination of 4 highlights results in an AUROC of 0.9757 and 0.9996 in testing and preparing datasets in a separate request.

In the wake of evaluating 253 clinical examples of blood from Wuhan, a few scientists have distinguished 11 critical related records that could uphold as separation gadgets of Coronavirus over the medical services capable of the speedy conclusion expressed in [16]. The commitments showed that 11 related files had been extricated after drawing in irregular timberland (RF) calculation with 95.95% of exactness and particularity of 96.97% in separate requests. Artificial intelligence and ML applications have been utilized to anticipate and assess what is happening. The clever methodology estimated and expected in 1-6 days of, generally speaking, ten patients in conditions of Brazil by using a stacking group through an SVM relapse calculation on collective positive instances of Coronavirus from Brazilian information. Consequently, it builds the transient expectation strategy for alarming the medical care capable and government for dealing with the pandemic as expressed in [17]. The contemporary examinations suggested a clever methodology by using managed recursive complex classifier known as XGBoost on mammographic and clinical boundary datasets. Later executing the strategy, the analysts recognized three conspicuous highlights from 75 clinical elements and tests of blood test results as 90% of precision in assessing and estimating the Coronavirus patient as broad, moderate, and severe as expressed in [18].

The gauge approach is connected with a choice rule for anticipating rapidly and gauges the contaminated individuals at the most extreme gamble; the patients who are pronounced as tainted ought to be considered for severe mindfulness and fundamentally diminish the brevity rate. The estimating approach in light of Canadian using time series has been created by connecting profound learning calculation over a broad momentary memory organization. The work [19] presents that few scientists have recognized a critical boundary and went for the gold course with an expectation of the finishing point of the current plague of SARS-CoV2 in Canada and the whole universe. The suggested approach expectation is finishing the end of the SARS-CoV-2 pandemic in Canada would be roughly June 2020. The work [20] presents that, contingent upon accumulated information from the College of John Hopkins, the assessment is logically exact as an original tainted case has been diminished quickly. The reasonable estimating strategy has been projected by coordinating decency of gauging model in light of wavelet and, further, the time series approach given auto-backward coordinated usually moving as portrayed in [21].

The goal of the contemporary commitments is the forecast the Coronavirus scope from the clinical reports of the objective patient. However, the PC helped clinical practices are fundamental to work on the awareness and particularity of the Coronavirus expectation; the seriousness (mortality extent) of the unhealthy patient forecast is critical to treat the patient with redid course of clinical suggestions. Worried about the target of mortality scope expectation, this original copy proposed a measurable investigation scale.

The contemporary commitment Slope Supporting Endurance Model (GBSM) [22] has investigated the presentation of broadened regulated learning strategies to foresee the mortality extent of an ailing person who tried positive in the Coronavirus test. Nonetheless, the main credits' age and sex has prescribed as ideal for preparing the classifier, and the characterization cycle is more driven to foresee an opportunity to release. The other commitment Individual-Level Casualty Forecast Model [23] that spurred from the release time expectation by Inclination Supporting Endurance Model [22] has tried to distinguish the extent of artificial brainpower strategies to foresee the singular casualty extent of the patients tested positively to the Coronavirus. These commitments utilize the segment highlights, which incorporate any persistent sickness as positive or negative. The other segment highlights connected with ailment have yet to be used in these contemporary commitments. Differentiation to these strategies, the proposed method has utilized the expanded elements connected with clinical finding results of the patients to foresee the Coronavirus contamination and its seriousness, a positive personal Coronavirus test.

III. MATERIALS AND METHODS

Chest x-rays, abbreviated CXR radiographs, of patients classified as high, medium, low, and healthy were used as critical input to the proposed model. The class label "high" indicates a high risk of mortality, "medium" indicates the need for ICU services often with mechanical ventilation, "low" indicates the need for home isolation with close monitoring of oxygen saturation's, and "healthy" indicates a negative test result. The proposed model begins with preprocessing the input radiographs in order to produce higher-quality source radiographs with less noise, which can be accomplished using any of the well-known image processing noise filters. The resulting image in this case contains residual effects from background subtraction. As a result, the preprocessing phase also includes segmentation of the source radiographs in order to differentiate the region of interest (ROI) from the surrounding areas of the source images. The final image would be a waveform with the least possible effect on the boundary. Additionally, the CXR ROI border with thick boarder is smoothed during the preprocessing phase.

3.1 The Features

To train the classifiers, CXRs were evaluated. In this case, the given CXR images will be processed to extract the region of interest, which is the infected lung's white mass. The feature engineering is applied to detect suggested features in a CXR radiograph. To train the chosen classifier, diverse characteristics listed as "Haralick texture features", "HOG (Histogram of Oriented Gradients) features", and "LBP (Local Binary Pattern) features" should be engineered further.

3.1.1 Haralick Texture Features

The texture of a picture may be measured using the co-occurrence matrix, which is dependent on the image's grayscale values or intensity, as well as multiple color dimensions. Because these matrices (of co-occurrence) are large and dispersed, several measures of the two dimensional matrix are investigated for a specific set of characteristics. The two matrices are turned at different angles (0 degree, 45 degree, 90 degree, as well as 135 degree) and various characteristics are assessed; these features are known as Haralick texture features and are retrieved from the enlarged picture. The grey-level co-occurrence matrix (GLCM) of dimension Ng, wherein Ng represents multiple intensity levels in

the given radiograph, which is a square matrix. The matrix element indexed at (i, j) is created by counting how many instances a pixel with index 'i' is next to a pixel at index 'J'. Further divides the complete matrix with the count of comparisons done. Each item represents the likelihood of a pixel with index 'i' being next to a pixel with index 'J'. The textural characteristics of Haralick are mentioned in the table below (table 1)

TABLE 1: THE LIST OF HARALICK FEATURES

Entropy of difference.	The difference between an image's entropy values
Correlation information measures 1 and 2	The expression of the variables' combined probability density distributions decreases the conventional correlation coefficient.
Correlation coefficient with the greatest value.	The interaction of pixels in a radiograph on a linear scale
Contrast.	The difference between the maximum and least pixel values is measured.
Second angular moment (ASM).	The metric gauges the homogeneity of a picture.
Variance.	The average of squared deviations from the image's mean
Total variance	The sum of the image's variance values
Entropy.	The quantity of data that should be used for a radiograph
Average of the sum	The aggregate of all the image's mean values
Variance in comparison	The difference between an image's variance values
Correlation.	The Grey levels of adjacent pixels in the picture have a linear relationship.
Moment of Inverse Difference (IDM).	The image's local uniformity
Total entropy.	The entire quantity of data that must be encoded in a picture

3.1.2 HOG Features

The Hoard (Histogram of Situated Slopes) is a worldwide picture descriptor that is generally utilized for object acknowledgment in IP (picture handling) as well as CV (PC vision). The technique counts the times a slope direction happens in a specific region of an image [8]. Regional object aspect and form inside an image are described by the dispersion of intensity- gradients or edge-information, which is a key notion underpinning the histogram of directed gradients descriptor. The input image is segmented into small linked sections denotes as cells, and a HOG is created for the pixels of each cell. The conjunction of the stated histograms is then used as the descriptor. Regional histograms can indeed be normalized for better accuracy by computing an approximation of the intensity throughout a wider region of the picture, known as a block, and then each of the cells within a block will be normalized by using the resultant value. Because of this normalization, the inversion of changes in light and shading is improved.

3.1.2.1 Computation of Gradients

Within the features of HOG, a localized 1-D histogram of edge directions is accumulated across the pixels of each of the cells by splitting the picture window into tiny spatial sections called cells. The gradient values are computed throughout the calculation. Applying the 1-D centered, the mask of point-discrete derivatives at either of the both horizontal as well as vertical axes is the most typical approach. This approach necessitates filtering the image's color or intensity data using the filtering kernels [1,0, -1] as well as [1, 0, -1].The filter "Gaussian-smoothing" was used to produce the gradients, which were then tested using one of many masks of discrete derivatives.

3.1.2.2 Binning by Orientation

The next stage in the HOG features is to calculate the cell histograms. Depending on the direction of the gradient component placed on it, each of the pixels calculates Histogram Channel's weighted-vote, and then, the votes are aggregated into orientation-bins across small spatial areas called cells. Cells might be radial or rectangular in shape. According to the gradient's sign, which is either "signed" or "unsigned", the bins are uniformly spaced between "0-degree and "180-degree or 360-degree". The matching bin has found for orientation of each of the pixels.

3.1.2.3 Block of Descriptors

The gradient intensities in the Features extra4ted must be localized. The cells must be normalized, which means they must be grouped together into

bigger, physically related blocks. The HOG description is the fusion of the elements and portions of each block region's cell histograms that are in normal form. These blocks frequently overlap, implying that each of these cells contributes to the final description more than once. There are two types of block geometries: the blocks of R-HOG that are in rectangle shape and the blocks of C-HOG that are in circle shape.

3.1.2.4 Normalization Block

Block normalization may be done in four distinct ways. Let V be the quasi vector representing histograms of a particular block, where ||VI|K is its k-norm for 1<=k<=e a tiny constant of approximate value. The normalizing value is then chosen from the list below:

$$L_2\text{norm} = v*\sqrt{(||\mathbf{v}||_2 + \mathbf{e}^2) - \mathbf{1}}$$

L2-hys: As in, L2-norm, clipping (v<= 0.2), and converts it again to normal form as follows

L1norm =v * (|v||$_1$+ e) -^{1}

$$\text{L1sqrt} = \sqrt{\mathbf{v} * (||\mathbf{v}||_2 + \mathbf{e}) - \mathbf{1}}$$

Furthermore, the L2-hys technique may be determined by considering the L2-norm, the clipped result, and performs normalizing again.

3.1.3 Features of LBP

LBP (Local-Binary Pattern) is the description of an image, which explores how that image appears in a limited area surrounding a pixel. The fundamental LBP-operator was based on the premise that texture delivers two distinct items locally: a pattern and the intensity of that pattern. The LBP-operator operates on a 3X3 block of pixels of the given image. To produce a tag for the central pixel, the block's pixels should be thresholded by the intensity of the central pixel, multiplied by squares, and then aggregated. Because the neighbourhood is made up of 8 pixels, there are a total of 28 = 256 different labels that may be created using the corresponding grey-values of the central pixel and the other pixels of the neighbourhood [9]. Along with global features, the average and deviation of LBP-region level features of an image are determined and utilized in classification.

3.2 Feature Optimization

The task of Feature optimization discovers the HOG features, Haralick texture features, and local binary patterns that are exclusive or more optimal towards the radiograph images of exclusive class label. The MWU-Test (Mann-Whitney U Test) [27] is one of the variety evaluation techniques that doesn't include a driven to the dispersion design that is expected in many datasets with different names. Since this test is a non-parametric test, it makes no assumptions about score distribution. However, various assumptions are made, such as the randomness of the observation selected from the population, the independence of the observations and bilateral independent, and the use of an arbitrary measuring tool. This non-parametric test is an alternative to an independent t-test. It's utilized to check whether two examples are from a similar populace or on the other hand if the offered viewpoints are bigger than different perceptions. Coming up next is a portrayal of the MWU-Test execution process.

The documentations v_1,v_2 indicate the vector dispersions utilized to contribute to the technique MWU-Test to finish up the extent of variety between comparing vectors, as per the following. At first, every one of the passages of vectors v_1,v_2 is moved to another vector. Further, sort the vector v in a climbing request of the qualities and let the files of the arranged upsides of the vector v compare positions R. The normal of the indistinguishable qualities' records will be the position of the relative multitude of separate indistinguishable rates. Further depiction signifies the places doled out to the vector's grades v_1 as a set R_1 and the places relegated to the vector's qualities v_2 as a set R_2. Later the cycle tracks down the total of the passages in the set R, as RS_1, which is additionally used to decide the position aggregate edge RST of the vector, as continues in (Eq 1):

$$\text{RST}_1 = \text{RS}_1 = \frac{1v1x(Iv11+1)}{2} \quad \text{.... (Eq 1)}$$

the notation lv, I denotes the size of the vector v_1.

Similarly, the rank-sum threshold RST, of the vector v_2 will be determined as follows

$$\text{RST}_1 = \text{RS2} = \frac{1v2x(Iv21+1)}{2} \quad \text{.... (Eq 2)}$$

In (Eq 2), notational v_2 I denotes the size of the vector v_2, and the notation RS, denotes the sum of the ranks of the entries in vector v_2 those listed in a set R_2.

Then the rank-sum threshold RST of the vectors' entries v_1,v_2 is the sum of rank-sum thresholds RST, RST_2 of the vectors v_1, v_2 followed in (Eq 3).

$RST_1 = RS_1 + RST2$ ... (Eq3)

Then find the z-score [28] as follows:

Initially, find the mean mrst and standard deviation dRST as follows in (Eq 4), (Eq 5):

$$\text{mRST} = \frac{\boldsymbol{RST}}{\mathbf{2}} \quad \text{... (Eq4)}$$

$$\text{dRST} = \sqrt{\frac{|V1|*v|+1)|}{|v|}} = \sqrt{\frac{|V1|*|v2|)|}{|v|}} [||v| + 1] - \sum_{il}^{k} \frac{t3-t1}{|v|*(|v|-1)}] \ldots Eq5)$$

The notation k signifies the number of separate rankings, and notation t, stands for the number of elements with the same rank.

Further, the z-core assesses as follows in (Eq 6)

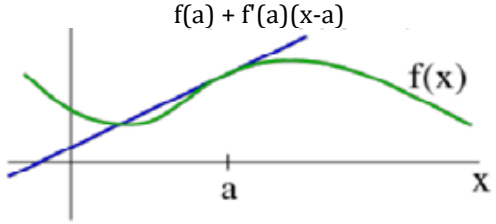
$$Z = \frac{RST - m_{RST}}{d_{RST}} \quad \text{... (Eq6)}$$

Then find the p-value of the depicted z score in the z-table [29]. The vectors v_1,v_2 are deemed to be diverse if the p-value is bigger than the provided probability threshold (typically 0.1, 0.05, or 0.01). Else the distribution of vectors are similar.

3.3 Classification Approach

XGBoost is a widely used and fast implementation of the Gradient-Boosted Trees-based classification technique[30], which is built on expression optimization via the approximation of certain loss-expressions as well as the use of multiple regularization strategies. At an iteration t, the objective-expression which is the conjunction of loss-expression plus regularization strategies) that intends to minimize is as follows:The XGBoost objective of fitness is recurrent expressions, which is of recurrent CART learners that cannot be maximized using typical Euclid space optimization approaches [31]. As an illustration, it can be stated that, the optimum linear estimate for f(x) at position 'a' is [32]:

f(a) + f'(a)(x-a)



Where:

$$f(t) \approx f(a)(x-a) + \frac{1}{2}f^{x}(a)(x-a)2$$

$$\mathcal{E}^{(t)} \approx \sum_{t-1}^{n}(t(y_{11}\, y^{(t-1)} - g_1 f_t(x_i) + \frac{1}{2}h_i f_1^2(x_i)] + \Omega\,(f_t)$$

XG Boost Objective Using Second-order Taylor approximation

$$g = \partial_g(x-1)t(y_{t1}\, y^{(t-1)} and\ h_i = \partial_g^2(t-1)(y_{11}\, y^{(t-1)})$$

The loss expression's Ist as well as 2nd ordered gradient stats

The loss expression's 1st and 2nd ordered gradient statistics Finally, removing the constant components yields the following concise reduction aim at active step t:

$$\mathcal{E}^{(t)} = \sum_{t-1}^{n}(g_{1f_1}(x_1) + \frac{1}{2}h_{1f_t^2}(x_1)| + \Omega(f_t)$$

XGBoost basic objective

The next objective is to achieve a learner that maximizes the loss-expression at present iteration t, because the former is a summation of the quadratic units of a single variable that can be reduced using well-known techniques.

$$argmin_z Gx + \frac{1}{2}Hx^2 = -\frac{G}{H}, H > 0\ min_z\, Gx + \frac{1}{2}Hx^2 = -\frac{1}{2}\frac{G2}{H}$$

Reduce a basic quadratic formula to its simplest form

It is necessary to apply the Taylor approximations so we must convert the initial fitness expression to the Euclidean space in order to employ conventional optimization methods.

Consider the simplest linear approximate solutions of the expression f(x) as follows:

$$f(x) \approx f(a)f'(a)(x-a)$$

$$\Delta X = f_t(x_i)$$

Initial expression is merely Ax. Using Taylor's principle, we can convert f(x) to a minimal expression of x around over a point 'a'. Before the Taylor approximation, the fitness expression f(x) was the summation of t CART trees, but now it is merely the step t (present tree). Discrete fitness expression is required. The predicted value (t-1) is ax, the new learner that need to be added at step t, and f(x) is the loss expression 1 this time. This allows us to use the different methods to optimize the Euclidean space by defining the loss expression as a simple expression of each independent learner newly added. As previously indicated, the estimate at previous step (t-1) is 'a,' and the new learner (x-a) that requires at step t. As a result of using the second- order Taylor approximation, we obtain:

As iteration t requires the development of a learner that achieves the highest potential loss reduction. Hence, there is a method for "assessing the efficacy of a tree structure q," and the scoring expression is as follows

$$\pounds^{(t)}(q) = -\frac{1}{2}\sum_{j=1}^{T} k\frac{(\sum_{iCij} gi)2}{\sum_{i\in ij,} h_{i+\lambda}} + \gamma T.$$

It is extremely difficult to "enumerate all potential tree architectures k=1 throughout the program and q" and hence locate the tree with greatest control of the loss.

Notably, the aforementioned "quality scoring expression" returns the smallest loss assigned to a given tree structure, meaning that the actual loss expression is evaluated using appropriate weights. Thus, a method exists for computing the optimal leaf weights for every tree topology.

In action, following is performed to grow the learner:
Begin with a single root (takes all of the training samples)

Recurrent over all characteristics and their associated values, evaluating each conceivable decrease in splitting loss:

Gain = (lossib – lossrb) – (lossib – lossrb)

Loss of left branch $loss_{ib}$,loss of right branch $loss_{rb}$

The positive gain shall be expected from the optimal splitting, which should be greater than the min splitting gain argument, else the branch will cease developing. The algorithm described above is referred to as the "Exact Greedy Algorithm" with O(n*m) complexity, wherein n denotes the count of training instances and m denotes the dimensionality of the characteristics.

Consider the following scenario of two-class classification as well as the logging loss expression:

$$yIn(p) + (1-y)In(1-p) where\ p = \frac{1}{(1+e^{-z})}$$

Two – class classification having Cross – Entropy loss expression,

wherein y is the actual label in the range 0 to 1 as well as p denotes the approximated probability.Notably, p (estimate or quasi-probability) has to be derived after the exponential expression is applied to the result of the GBT prototype x. The model's output x is the aggregate of the learners of the CART-tree.

Thus, to reduce the log-loss optimization approach, it must first determine its first and second derivative (gradient as well as hessian) in regard to x. You may get gradient = (p-y) as well as hessian = p* at thread (1-p). Learners will aim to limit the log loss goal, as well as the leaf level scores, which are actual weights that reflects the significance as a sum across all available trees in the model are always altered to reduce the loss, by synthesizing the GBT approach, which is an aggregate of CART scores. As a consequence, the sigmoid activation expression should be applied to the output of GBT modelling techniques in the form of two-class classification likelihood score.

IV. Result and Discussion

4.1 The Data

The input X-rays with label (Positive or negative) has adopted from the renowned dataset called BIMCV-COVID19+ [33]. The chest x-rays have intended to cover a wide spectrum of thoracic entities, which is a significant escalation of the dataset set quality that compare to majority of existing datasets. The other significance of the dataset is the high resolution of the chest x-ray images. The first version of the dataset contains 1380 chest x-rays. Performance has been assessed under diversified metrics like overall prediction accuracy, sensitivity that denotes True-positive-rate (TPR), specificity that denotes True-negative-rate (TNR)[34], and precision scaled under diversified optimal features selected under different distance thresholds. The suggested method "SC2SSP" has been critically assessed by comparing it with the performance of the contemporary approaches RES-NET50[35] and Transfer Learning[36].

4.2 Performance Analysis

The cross validation has performed in four-fold, and the cross-validation metrics has been estimated for four labels "server, moderate, mild and benign". The metrics used to scale the significance of the model proposed and the models found in contemporary contributions RES- NET50 [25], Transfer-Learning [26] are precision (often referred as positive predictive value), recall, which also termed as

sensitivity, specificity that also termed often as true negative rate, decision accuracy, which is the ratio of specificity and sensitivity towards given total records of diversified labels, and harmonic mean, which indicates the F-Score of the predicted labels. Since the cross validation is performed on multiple labels, the metrics "micro precision, micro recall, micro specificity, micro specificity, and decision accuracy" have also being assessed.

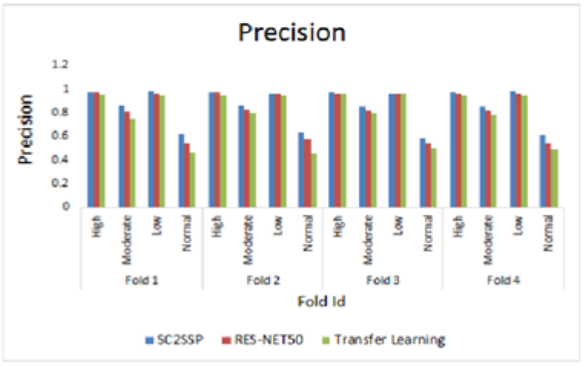


Figure 1: Graphical representation of comparison of proposed model SC2SSP and contemporary models RES-NET50 and Transfer-Learning in terms of metric precision over four folds.

In figure 1, graph has been plotted, where precision is represented on y-axis and 4 folds are represented on x-axis containing diversified labels such as high, medium, low, and healthy. These labels are compared among proposed model SC2SSP and contemporary RES-NET50 and Transfer-Learning. Thus, the suggested model beats existing models in all labels in terms of precision.

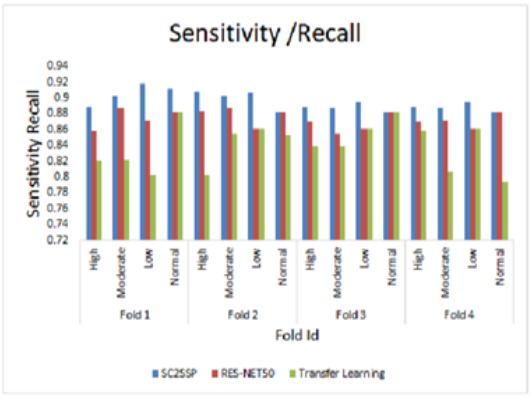


Figure 2: Graphical representation of comparison of proposed model SC2SSP and contemporary models RES-NET50 and Transfer-Learning in terms of metric sensitivity over four folds.

The ratio of genuine positives discovered to overall positives discovered is known as sensitivity. The recall measure used to compare the proposed SC2SSP model to the current RES-NET50 and Transfer-Learning models is shown in Figure 2. Three models with four labels are compared in the graph: high, medium, low, and healthy. As a result, the suggested model beats current models across all labels in terms of sensitivity.

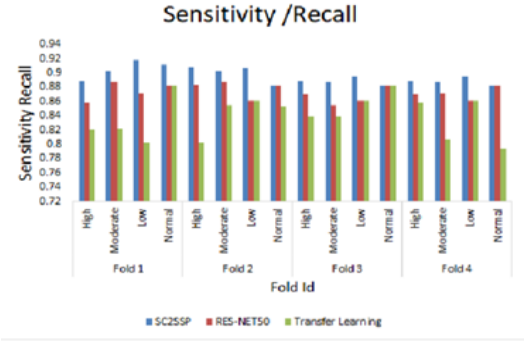


Figure 3: Metric specificity for the suggested SC2SSP and existing approaches RES- NET50, and Transfer-Learning models.

One of the essential metrics, the ratio of genuine negatives to the total number of true negatives assessed, is depicted in Figure 3. For the corpus of requirement specifications, the suggested model SC2SSP beats the other important models RES-NET50 and Transfer- Learning. Three models with four labels are compared in the graph: high, medium, low, and healthy. In terms of specificity, the suggested model beats current models across all labels.

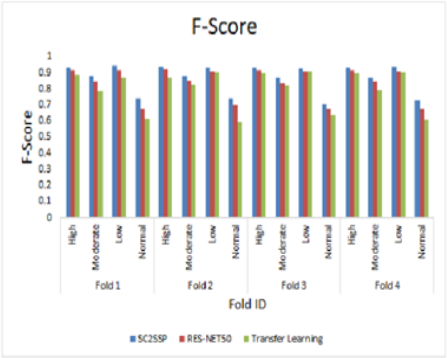


Figure 3: Graphical representation of comparison of proposed model SC2SSP and contemporary models RES-NET50 and Transfer-Learning in terms of metric F-score over four folds.

In figure 4, graph has been plotted, where F-score is represented on y-axis and 4 folds are represented on x-axis containing diversified labels such as high, medium, low, and healthy. These labels are compared among proposed model SC2SSP and contemporary models RES- NET50 and Transfer-Learning. Hence, F-score denotes that the proposed model SC2S SP outperforming the contemporary models.

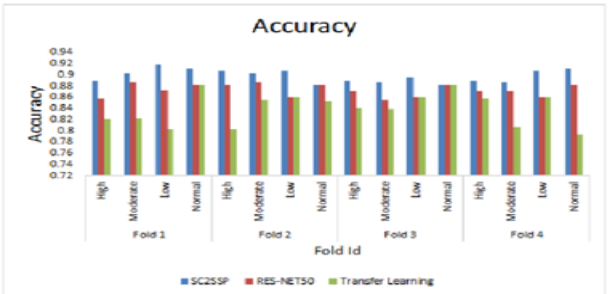


Figure 4: Graphical representation of comparison of SC2SSP and RES-NET50 and Transfer- Learning in terms of metric Accuracy over four folds.

The discovered values of the metric "Accuracy" of proposed and contemporary models over the 4 folds as demonstrated in figure 5. A contrast of suggested and existing models is presented, with each model denoting the severity of the disease as high, medium, low, or healthy. The forecast SC2SSP's accuracy in all labels outperforms current models.

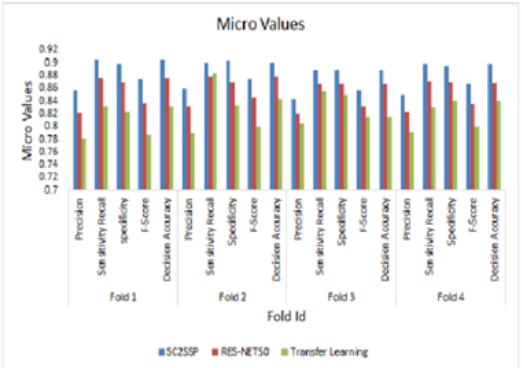


Figure 5: Graphical representation of comparison of all the metrics representing suggested model SC2SSP and existing methods RES-NET50 and Transfer-Learning

In figure 6, the graph representing the micro values of the metrics "precision", "sensitivity", "specificity", "F-score" and decision accuracy of SC2SSP and contemporary models RES- NET50 and Transfer-Learning. From above statistics portrayed in the figure, the performance of proposed model in all the metrics performs more superior over the contemporary models.

V. CONCLUSION

The study presents a multiclass supervised learning approach for predicting the severity and scope of SAR-COV2 infections. The results of the study show that this approach can effectively forecast the impact of SAR-COV2.This is an important contribution to the field as it helps to address one of the critical challenges in the fight against the pandemic - the lack of accurate and timely predictions of the spread and impact of the virus. The approach used in this study leverages machine learning algorithms to analyze large amounts of data, allowing for more comprehensive and accurate predictions. The results of this study have significant implications for public health decision-making, as they can inform the development of more effective intervention strategies and the allocation of resources.In addition to its practical applications, this study also highlights the importance of interdisciplinary collaborations between researchers in computer science, mathematics, and public health. The integration of different perspectives and expertise is crucial in addressing complex global health challenges like the SAR-COV2 pandemic.

AUTHOR CONTRIBUTIONS

Shaik Khasim Saheb B. Narayanan and Thota Venkat Narayana Rao conducted the research, analyzed the data and wrote the paper then all authors had approved the final version.

REFERENCES

1. Aswathy A.L.;Anand Hareendran S.;Vinod Chandra S.S.; (2021). COVID-19 diagnosis and severity detection from CT-images using transfer learning and back propagation neural network . Journal of Infection and Public Health, p1-11.
2. M. Turkoglu; (2021). COVID-19 Detection System Using Chest CT Images and Multiple Kernels-Extreme Learning Machine Based on Deep Neural Network . IRBM, p1-8.
3. Ahuja, Sakshi; Panigrahi, Bijaya Ketan; Dey, Nilanjan; Rajinikanth, Venkatesan; Gandhi, Tapan Kumar (2020). Deep transfer learning-based automated detection of COVID-19 from lung CT scan slices. Applied Intelligence, p1-15.
4. Yazan Qiblawey;Anas Tahir;Muhammad E. H. Chowdhury;Amith Khandakar;Serkan Kiranyaz;Tawsifur Rahman;Nabil Ibtehaz;Sakib Mahmud;Somaya Al Maadeed;Farayi Musharavati;Mohamed Arselene Ayari; (2021). Detection and Severity Classification of COVID-19 in CT Images Using Deep Learning . Diagnostics, p1-19.
5. Daryl L. X. Fung;Qian Liu;Judah Zammit;Carson Kai-Sang Leung;Pingzhao Hu; (2021). Self-supervised deep learning model for COVID-19 lung CT image segmentation highlighting putative causal relationship among age, underlying disease and COVID-19 . Journal of Translational Medicine, p1-18.
6. Sekeroglu, Boran; Ozsahin, Ilker (2020). Detection of COVID-19 from Chest X-Ray Images Using Convolutional Neural Networks. SLAS TECHNOLOGY: Translating Life Sciences Innovation, p1-13.
7. Seyed Masoud Rezaeijo;Mohammadreza Ghorvei;Razzagh Abedi-Firouzjah;Hesam Mojtahedi;Hossein Entezari Zarch; (2021). Detecting COVID-19 in chest images based on deep transfer learning and machine learning algorithms . Egyptian Journal of Radiology and Nuclear Medicine, p1-12.
8. Yarong Li1,a , Yizhang Jiang1, 2,b, Yi Gu1, 2,c, and Pengjiang Qian1. (2022). An Automatic Detection Method for COVID-19 in CT Images. Journal of Physics: Conference Series, p1-9.
9. Mohamed Ramzy Ibrahim;Sherin M. Youssef;Karma M. Fathalla; (2021). Abnormality detection and intelligent severity assessment of human chest computed tomography scans using deep learning: a case study on SARS-COV-2 assessment . Journal of Ambient Intelligence and Humanized Computing, p1-24.
10. Mehmet Akif Cifc. (2020). Deep Learning Model for Diagnosis of Corona Virus Disease from CT Images. International Journal of Scientific & Engineering Research. 11 (4), p1-6.
11. Emrah Irmak. (2021). COVID-19 disease severity assessment using CNN model. IET Image Processing, p1-11.
12. Tawsifur Rahman1 , Alex Akinbi2* , Muhammad E. H. Chowdhury1 , Tarik A. Rashid3 , Abdulkadir Şengür4 , Amith Khandakar1 , Khandaker Reajul Islam1 and Aras M. Ismael5. (2022). COV-ECGNET: COVID-19 detection using ECG trace images with deep convolutional neural network. Health Information Science and Systems, p1-16.
13. HANAN S. ALGHAMDI 1 , GHADA AMOUDI1 , SALMA ELHAG1 , KAWTHER SAEEDI 1 , AND JOMANAH NASSER2. (2020). Deep Learning for Automated Recognition of Covid-19 from Chest

X-ray Images. IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY SECTION. 9, p1-20.

14. DEVRİM AKGÜN ABDULLAH TALHA KABAKUŞ ZEHRA KARAPINAR ŞENTÜRK ARAFAT ŞENTÜRK ENVER KÜÇÜKKÜLAHLI . (2021). A transfer learning-based deep learning approach for automated COVID-19diagnosis with audio data . Turkish Journal of Electrical Engineering & Computer Sciences. 29 (8), p 2807 – 2823.
15. Subhagata Chattopadhyay. (2021). Towards Grading Chest X-rays of COVID-19 Patients Using A Dynamic Radial Basis Function Network Classifier . Artificial Intelligence Evolution. 2 (2), p1-15.
16. Tayarani-N., Mohammad-H. (2020). Applications of Artificial Intelligence in Battling Against Covid-19: A Literature Review. Chaos, Solitons & Fractals, p1-62.
17. Somil Vasal, Sourabh Kumar Jain and Ashok Verma. (2020). COVID-AI: An Artificial Intelligence System to Diagnose COVID-19 Disease. International Journal of Engineering Research & Technology (IJERT). 9 (8), p62-67.
18. Yassine Bouchareb, Pegah Moradi Khaniabadi, Faiza Al Kindi, Humoud Al Dhuhli, Isaac Shiri, Habib Zaidi and Arman Rahmim. (2021). Artificial intelligence-driven assessment of radiological images for COVID-19 . Computers in Biology and Medicine . 136, p1-18.
19. PIR MASOOM SHAH1,6, FAIZAN ULLAH1 , DILAWAR SHAH1 , ABDULLAH GANI (SENIOR MEMBER, IEEE)2,3, CARSTEN MAPLE (MEMBER, IEEE)4,5, YULIN WANG6 , SHAHID1 , MOHAMMAD ABRAR7 , SAIF UL ISLAM. (2021). Deep GRU-CNN model for COVID-19 detection from chest X-rays data. IEEE. 9, p1-13.
20. Sangjoon Park, Gwanghyun Kim, Jeongsol Kim1 , Boah Kim and Jong Chul Ye. (2021). Federated Split Vision Transformer for COVID-19 CXR Diagnosis using Task-Agnostic Training. Conference on Neural Information Processing Systems, p1-14.
21. Amar Kumar Verma;Inturi Vamsi;Prerna Saurabh;Radhika Sudha;Sabareesh G.R.;Rajkumar S.; (2021). Wavelet and deep learning-based detection of SARS-nCoV from thoracic X-ray images for rapid and efficient testing . Expert Systems with Applications, p1-16.
22. Li X, Geng M, Peng Y, Meng L, Lu S. Molecular immune pathogenesis and diag- nosis of COVID-19. J Pharm Anal 2020.
23. Wang, Linda; Lin, Zhong Qiu; Wong, Alexander (2020). COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Scientific Reports, 10(1), p1-12.
24. Borakati, Aditya; Perera, Adrian; Johnson, James; Sood, Tara (2020). Diagnostic accuracy of X-ray versus CT in COVID-19: a propensity-matched database study. BMJ Open, 10(11), p1-14.
25. Loey, Mohamed; Manogaran, Gunasekaran; Khalifa, Nour Eldeen M. (2020). A deep transfer learning model with classical data augmentation and CGAN to detect COVID-19 from chest CT radiography digital images. Neural Computing and Applications, p1-13.
26. Zheng, Zhaohai; Peng, Fang; Xu, Buyun; Zhao, Jingjing; Liu, Huahua; Peng, Jiahao; Li, Qingsong; Jiang, Chongfu; Zhou, Yan; Liu, Shuqing; Ye, Chunji; Zhang, Peng; Xing, Yangbo; Guo, Hangyuan; Tang, Weiliang (2020). Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. Journal of Infection, p1-37.
27. Wang, Xinggang; Deng, Xianbo; Fu, Qing; Zhou, Qiang; Feng, Jiapei; Ma, Hui; Liu, Wenyu; Zheng, Chuansheng (2020). A Weakly-supervised Framework for COVID-19 Classification and Lesion Localization from Chest CT. IEEE Transactions on Medical Imaging, (), 1–11.
28. Pathak, Y.; Shukla, P.K.; Tiwari, A.; Stalin, S.; Singh, S.; Shukla, P.K. (2020). Deep Transfer Learning Based Classification Model for COVID-19 Disease. IRBM, p1-16.
29. Han, Z., Wei, B., Hong, Y., Li, T., Cong, J., Zhu, X., … Zhang, W. (2020). Accurate Screening of COVID-19 Using Attention-Based Deep 3D Multiple Instance Learning. IEEE Transactions on Medical Imaging, 39(8), p2584–2594.
30. Ozsahin, Ilker; Sekeroglu, Boran; Musa, Musa Sani; Mustapha, Mubarak Taiwo; Uzun Ozsahin, Dilber (2020). Review on Diagnosis of COVID-19 from Chest CT Images Using Artificial Intelligence. Computational and Mathematical Methods in Medicine, 2020, p1–10.
31. Jin, Cheng; Chen, Weixiang; Cao, Yukun; Xu, Zhanwei; Tan, Zimeng; Zhang, Xin; Deng, Lei; Zheng, Chuansheng; Zhou, Jie; Shi, Heshui; Feng, Jianjiang (2020). Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. Nature Communications, 11(1), p1-14.
32. Ahuja, Sakshi; Panigrahi, Bijaya Ketan; Dey, Nilanjan; Rajinikanth, Venkatesan; Gandhi, Tapan Kumar (2020). Deep transfer learning-based automated detection of COVID-19 from lung CT scan slices. Applied Intelligence, p1-15.
33. Xiao, Lu-shan; Li, Pu; Sun, Fenglong; Zhang, Yanpei; Xu, Chenghai; Zhu, Hongbo; Cai, Feng-Qin; He, Yu-Lin; Zhang, Wen-Feng; Ma, Si-Cong; Hu, Chenyi; Gong, Mengchun; Liu, Li; Shi, Wenzhao; Zhu, Hong (2020). Development and Validation of a Deep Learning-Based Model Using Computed Tomography Imaging for Predicting Disease Severity of Coronavirus Disease 2019. Frontiers in Bioengineering and Biotechnology, 8, p1-11.
34. Pu, Jiantao; Leader, Joseph K.; Bandos, Andriy; Ke, Shi; Wang, Jing; Shi, Junli; Du, Pang; Guo, Youmin; Wenzel, Sally E.; Fuhrman, Carl R.; Wilson, David O.; Sciurba, Frank C.; Jin, Chenwang (2020). Automated quantification of

COVID-19 severity and progression using chest CT images. European Radiology, p1-11.
35. Shen, Cong; Yu, Nan; Cai, Shubo; Zhou, Jie; Sheng, Jiexin; Liu, Kang; Zhou, Heping; Guo, Youmin; Niu, Gang (2020). Quantitative computed tomography analysis for stratifying the severity of Coronavirus Disease 2019. Journal of Pharmaceutical Analysis, p1-28.
36. Shan, Fei; Gao, Yaozong; Wang, Jun; Shi, Weiya; Shi, Nannan; Han, Miaofei; Xue, Zhong; Shen, Dinggang; Shi, Yuxin (2020). Abnormal Lung Quantification in Chest CT Images of COVID19 Patients with Deep Learning and its Application to Severity Prediction. Medical Physics, p1-49.

Shaik Khasim Saheb received the B.Tech degree in computer science and engineering from JNTUK University, Andhra Pradesh, India, in 2010 and the M.Tech degree in computer science and engineering from VIT Univeristy, Tamilnadu, India in 2014 and Pursuing Ph.D in computer science and engineering from Annamalai University, Chidambaram, and Tamilnadu, India respectively. Currently, he is an Assistant Professor at the Department of Computer Science and Engineering, Sreenidhi Institute of Sciene and Technology, Hyderabad, India. His research interests include digital image processing, machine learning and deep learning. He can be contacted at email: shaikkhasims@sreenidhi.edu.in

B. Narayanan, received B.E. M.E., and Ph.D. degrees in Computer Science and Engineering from Annamalai Univeristy, Chidambaram, Tamilandu, India in 2005, 2007 and 2017 respectively. He is an Assistant Professor in Computer Science and Engineering wing, Annamalai Univerisity, Chidambaram, India since 2008. He is having overall 14 years of academic teaching experience. He has received many awards in teaching as well as resrach. He has also published various resach papers in peer reputed international journals. He is guide to many research scolars. His research interest includes Data Mining, Big Data analytics, Machine Learning, Deep Learnining, Artificail Intelligence and Digital Image Procesing. He can be avaialbe at email: narayanan.bk@gmail.com

Thota Venkat Narayana Rao, received B.E. and M.Tech. degrees in Computer Science and Engineering from JNTUH, Hyderabad. India 1991 and 2009, respectively, and the Ph.D. degree in Computer Science and Engineering from JNTUK , Kakinada , India. He has been a Professor of CSE and HoD-CSE IOT in sreenidhi Institute of Science and Technology, Hyderabad , India. He is putting a experience of 30 years in both academia and industry. He has authored or coauthored more than 190 refereed journal and conference papers, 10 book chapters, with IEEE, Elsevier and Springer etc. His research interests include Medical Image Processing , Software Engineering, Cyber Security and Disruptive Technologies. He has many Awards and Patents to his credit . He can be contacted at email: venkatnarayayanaraot@sreenidhi.edu.in