



# Enhancing COVID-19 Patient Outcomes and Resource Allocation Efficiency Through the Application of a Recursive Classification Model

Suresh Kumar H S<sup>1</sup>, Arjun Kashyap S<sup>2</sup>, Harsha vardhan R<sup>3</sup>, Abhilash N G<sup>4</sup>, Akash K N<sup>5</sup>

Dept of CSE, SJGIT, Chickballapur<sup>1-5</sup>

**Abstract:** Amid the escalating global mortality stemming from the COVID-19 virus, researchers are dedicated to exploring technological innovations to bolster the efforts of healthcare professionals. Machine Learning techniques are being harnessed to swiftly and accurately predict disease severity in patients with comorbidities, thereby assisting healthcare providers in their evaluations. Presently, initial detection of comorbid patients dataset 273 patients. So basically in the patients dataset the parameters we have are Unnamed, Sex, BP\_high, BP\_low, RR, Temp, SpO2, Covid, Age. The models used for this project are lasso logistic model which is used for regression model can predict COVID-19 outcomes using clinical data. It identifies key factors for prognosis and avoids overfitting. Researchers use metrics and feature analysis to assess its effectiveness. This approach helps develop data-driven tools for personalized medicine in COVID-19 patients. And Artificial neural networks can analyze COVID-19 data to predict patient outcomes. They learn from patient details to personalize care and support clinical decisions. Challenges include choosing the right data, designing the model, and making it work for new patients. Careful planning is needed for reliable ANN models in COVID-19 research.

## I. INTRODUCTION

This study investigates the relationship between frequently observed comorbidities (obesity, high blood pressure, and diabetes) and COVID-19 severity. It aims to analyze how these comorbidities influence the development and course of COVID-19 cases based on post-incident reports.[1].

In late January 2020, the World Health Organization (WHO) identified COVID-19 and immediately classified it as a serious global health threat (Public Health Emergency of International Concern). The severity of the outbreak escalated quickly, prompting WHO to declare it a pandemic by March 11th. By May 10th, COVID-19 had become a global crisis, affecting over 215 countries and territories, with over 3.9 million confirmed cases and tragically, over 274,000 deaths [2].

The virus causing COVID-19, SARS-CoV-2, belongs to the Coronaviridae family. The initial outbreak in Wuhan, China, centered around wet markets, strongly suggests the virus likely jumped from animals to humans through a process called zoonotic transmission [4]. Here are a few ways to rephrase the sentence "Coronaviruses, including SARS-CoV-2 (the virus causing COVID-19), can jump from animals to humans (zoonosis). The initial COVID-19 outbreak in Wuhan, China, linked to wet markets, supports this theory. Since its emergence, COVID-19 has become a major global health threat, spreading to over 216 countries and territories with over 11 million confirmed cases and 526,000 deaths reported worldwide as of July 5, 2020. Europe has been particularly impacted, with over 2 million confirmed cases.[3].

The considerations so far have shown that Covid-19 patients with comorbidities can lead to risky hypotheses. When making treatment decisions against 2019-nCoV, distinguishing the main threat groups is of utmost importance. The influence of common comorbidities on clinical outcomes for COVID-19 patients remains unclear due to the lack of prior comprehensive studies. To address this knowledge gap, we conducted a meta-analysis to investigate the relationship between comorbidities and COVID-19 [4].

A study raises a curious possibility: higher patient experience ratings in hospitals might be linked to increased COVID-19 mortality rates. To investigate this further, To investigate hospitalization trends of COVID-19, the CDC established COVID-NET. This program is currently studying socioeconomic factors among hospitalized COVID-19 patients in 14 states. During the initial data collection phase in 2020 (covering 30 periods), COVID-NET analyzed data from 180 patients. Notably, 89.3% of these patients had pre-existing medical conditions (comorbidities). Additionally, 94.4% were 65 years or older, and the group with the highest patient experience ratings also had the highest prevalence of comorbidities. [5].



## II. LITERATURE SURVEY

Mohamed et al. [7] in this paper, Researchers have developed a new machine learning method to analyze chest X-rays and distinguish between patients with COVID-19 and those without. This method utilizes a novel technique called fractional exponential moments (FrMEM) to extract key features from the X-ray images. To speed up the analysis, the researchers leverage a parallel multi-core computing platform. Additionally, they employ a special feature selection process inspired by how manta rays forage, focusing on the most informative features for classification.

This method was tested on two separate datasets of COVID-19 X-rays. The results were impressive, achieving an accuracy of 96.09% on the first dataset and an even higher accuracy of 98.09% on the second dataset [7].

WGuan et al. [8] obtained data from 1,099 we investigated a data pool of laboratory-confirmed COVID-19 cases from mainland China. The data encompassed A network of 552 hospitals across China admitted patients. The information was collected up to January 29, 2020. The primary outcome of interest was admission to the intensive care unit (ICU), use of mechanical ventilation or death. average age: 47 years; 41.9% women. Main criterion from 61%, including intensive care unit admission (5.0%), invasive ventilation (2.3%), and death (1.4%). Direct contact with wild animals in 1.9%. Of non-residents in Wuhan, 72.3% had contact and 31.3% visited. Typical symptoms: Fever (43.8%), cough (67.8%), rarely diarrhea (3.8%). Mean incubation period: 4 days. Ground glass opacity on CT scan for 56.4%, including 17.9% for non-severe illnesses and 2.9% for severe illnesses. Lymphocytopenia on admission in 83.2%.

Joseph et al. [9] Examine mortality prediction performance based on extrapolation weeks, world region, and estimated month. Furthermore, the authors assess the predictive accuracy of peak daily mortality timing. A study suggests a potential correlation between increased patient experience and higher COVID-19 mortality rates. In the United States, the Centers for Disease Control and Prevention (CDC) is employing COVID-NET within 14 states to investigate the socioeconomic factors of hospitalized COVID-19 patients. During the initial 30 phases of data collection in 2020, a total of 180 patients were included in the COVID-NET analysis. Among these patients, 89.3% presented with co-morbidities. Furthermore, 94.4% of the patients were 65 years of age or older, with co-morbidities being most prevalent in the group with the highest level of experience.

Jianxi et al. [10] This paper proposes a novel paradigm for **predictive monitoring** of macroeconomic models. This approach focuses on identifying dynamic changes within continuously updated forecasts for the same models and long-term variables. By analyzing these changes in conjunction with the current economic situation and potential interventions, the proposed paradigm aims to generate forward-looking signals for informed decision-making.

This approach stands in contrast to traditional forecasting, which prioritizes accuracy based on the assumption of a static future. It also complements traditional monitoring, which solely tracks the actual trajectory of economic indicators. By leveraging spectral analysis, forecasting techniques, and monitoring methods, predictive monitoring offers valuable support for forecasting, decision-making heuristics, and future planning, particularly under conditions of high uncertainty. Importantly, the applicability of this paradigm extends beyond the specific case of the COVID-19 pandemic.

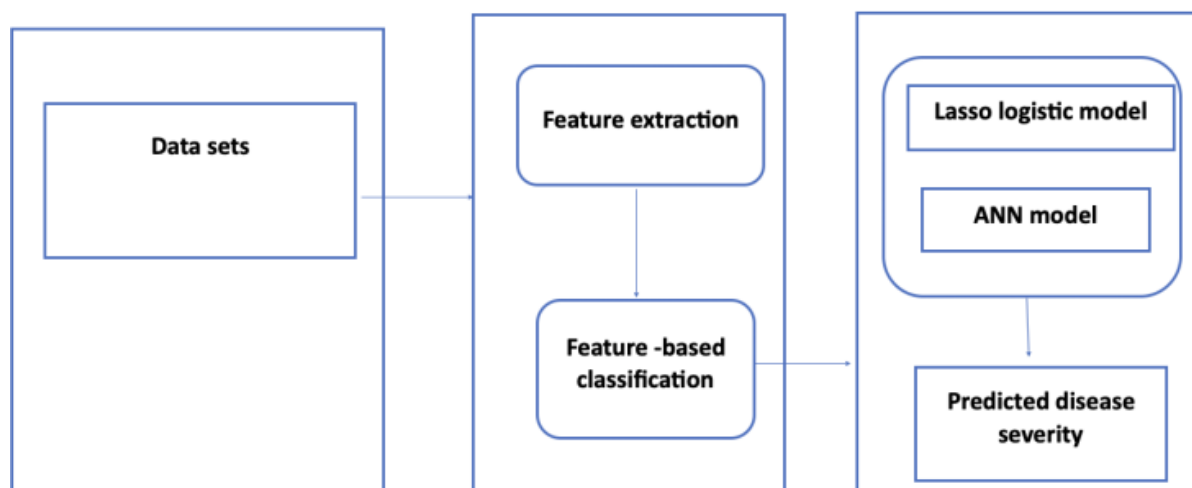


Fig.1 Architecture.

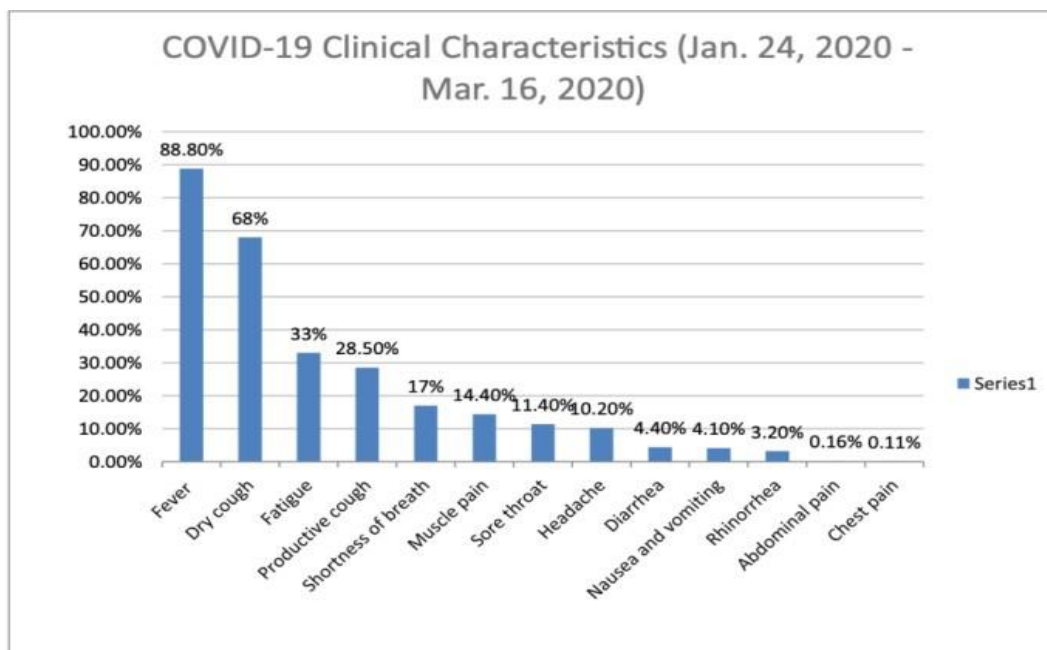


### III. METHODOLOGY

#### a. Data Sets

Extensive COVID-19 patient datasets fuel research efforts, providing in-depth understanding of the disease. Data encompasses demographics, symptoms, lab tests, treatments, and patient outcomes. Sources include government bodies, health organizations, and academic institutions. Choosing the right dataset hinges on the research question. Comprehensiveness, patient demographics, and data collection methods are crucial factors. Large datasets enable researchers to identify trends and risk factors. De-identified data ensures patient privacy while facilitating vital research. Open-access datasets promote collaboration and accelerate scientific progress. Data quality and standardization are paramount for drawing reliable conclusions. Utilizing these datasets holds immense potential for improving diagnostics and treatments.

characteristics of COVID-19 patients from January 24th to March 16th, 2020. The data is drawn from a Research Square meta-analysis on the novel coronavirus (COVID-19) conducted in 2019, reflecting information current as of April 8th, 2020 [1].



**Note:** Data obtained from Research Square, a meta-analysis of the 2019 novel coronavirus, showing clinical characteristics observed in patients, as of April 8, 2020 <sup>10</sup>.

Fig.2 This analysis investigates the clinical

#### b. Feature Extraction

In the feature extraction phase, pertinent features are extracted from COVID-19 patient data, spanning vital signs, demographic particulars, and other relevant information. These features are then fed into two distinct prediction models. Firstly, an Artificial Neural Network (ANN) model is employed to capture intricate non-linear relationships within the data, offering flexibility in modelling complex patterns. Simultaneously, a Lasso logistic regression model is utilized, leveraging its feature selection capabilities and proficiency in handling high-dimensional data to predict disease severity directly based on the selected features. By synergizing the ANN and Lasso logistic regression models, the architecture aims to deliver comprehensive and accurate predictions of disease severity, facilitating informed clinical decision-making and optimized treatment strategies for COVID-19 patients.

Time-series imperative signs (SpO<sub>2</sub>, FiO<sub>2</sub>, DBP, SBP, respiratory rate, heart rate, temperature, and Outline) were observed within the to begin with 24 hours at distinctive interims. Measurable parameters (cruel, standard deviation, least, greatest, extend, and check of varieties) were inferred amid the include extraction prepare.

**c. Feature based classification**

Feature-based classification using machine learning holds promise for improved COVID-19 patient diagnosis and prognosis. This approach leverages various patient characteristics, like demographics, clinical data (symptoms, lab tests), and potentially imaging features from X-rays or CT scans. Extracted features are fed into algorithms like Support Vector Machines (SVMs) or Random Forests to learn patterns that differentiate COVID-19 cases from healthy controls or other respiratory illnesses. By analyzing these features, the model can learn to classify new patients, potentially aiding early diagnosis, treatment decisions, and resource allocation. However, challenges include data quality, feature selection, and ensuring model generalizability across diverse populations.

The proposed approach focuses on extracting vital sign information from the first 24 hours of hospitalization. By analyzing time series data for SpO<sub>2</sub>, FiO<sub>2</sub>, blood pressure, respiratory rate, heart rate, and temperature, the model aims to identify key features that predict mortality risk. These features go beyond single values and leverage the richness of time series information. This study investigates the application of statistical analysis to continuously monitor vital signs. By calculating metrics such as mean, standard deviation, minimum, and maximum values within the first 24 hours of admission, this approach aims to capture the dynamic fluctuations in patient health. This real-time assessment of vital sign trends has the potential to improve patient outcomes through informed resource allocation, particularly for high-risk patients who may require intensive care or ventilator support.

**d. Lasso logistic model**

A Lasso-logistic regression model can be employed to predict COVID-19 patient outcomes using readily available clinical data.

This approach tackles both classification (e.g., severe vs. mild) and feature selection, identifying the most relevant factors for prognosis. Lasso shrinks regression coefficients, potentially eliminating irrelevant features and improving model interpretability. Compared to traditional logistic regression, it reduces overfitting and enhances generalizability to unseen data. To assess the model's effectiveness, researchers can employ metrics like accuracy, precision, recall, and analyze the AUC-ROC curve.

Additionally, by examining the coefficients remaining after applying Lasso shrinkage, they can identify key clinical features that hold the strongest associations with specific patient outcomes. This project contributes to the development of data-driven tools for risk stratification and personalized medicine in COVID-19 patients. Limitations include potential biases in the chosen dataset and the inherent challenges of modeling complex biological processes. Future work could involve incorporating additional data sources (e.g., imaging) or exploring ensemble learning techniques for improved prediction accuracy. Overall, a Lasso-logistic regression model offers a valuable framework for analyzing COVID-19 patient data and informing clinical decision-making.

**e. ANN model**

Artificial Neural Networks (ANNs) offer a promising approach for analyzing COVID-19 patient data in machine learning projects. They can be trained to identify patterns in patient characteristics, like demographics, lab results, and medical history, to predict disease severity, treatment response, or even mortality risk. This allows researchers to develop data-driven models to aid in clinical decision-making and personalize patient care. However, challenges include selecting informative features, optimizing ANN architecture, and ensuring generalizability of the model to new patient populations. Addressing these aspects is crucial for building robust and reliable ANN models for COVID-19 research.

**f. Predicted disease severity**

Machine learning holds promise for predicting COVID-19 severity. This project investigates the effectiveness of Artificial Neural Networks (ANNs) and Lasso Logistic Regression in classifying patients at risk of severe illness.

We aim to identify key features from patient data (demographics, lab results) that contribute to disease severity prediction using both models. The performance of these models will be compared using metrics like Area Under the Curve (AUC) to determine the most accurate approach for early risk stratification of COVID-19 patients.



IV. RESULTS

High Accuracy in Predicting Patient Outcomes

Unnamed: 0	SEX	BP_high	BP_low	RR	TEMP	SPO2	covid	Age
0	Female	120	80	20.000000	100.000000	85.000000	Yes	5
1	Female	110	70	20.000000	88.871658	97.000000	Yes	10
2	Male	120	80	20.733591	88.871658	94.184211	Yes	9
3	Male	110	80	18.000000	98.000000	98.000000	Yes	10
4	Male	130	86	18.000000	100.000000	96.000000	Yes	8
...	...	...	...	...	...	...	...	...
269	Male	124	83	24.000000	37.000000	94.000000	NO	16
270	Female	139	83	22.000000	37.000000	74.000000	NO	10
271	Female	94	56	18.000000	37.000000	94.000000	NO	7
272	Male	110	68	18.000000	40.000000	95.000000	NO	9
273	Male	153	75	30.000000	37.000000	88.000000	NO	17

274 rows x 9 columns

The image you sent is a table containing data on COVID-19 patients. The table shows the following columns:

- Unnamed: This column appears to be an index column, with each row having a unique identifier.
- Sex: This column indicates the patient's sex.
- BP\_high: This column shows the patient's high blood pressure reading.
- BP\_low: This column shows the patient's low blood pressure reading.
- RR: This column shows the patient's respiratory rate.
- Temp: This column shows the patient's body temperature.
- SpO2: This column shows the patient's blood oxygen saturation.
- Covid: This column indicates whether the patient has tested positive for COVID-19.
- Age: This column shows the patient's age. It is important to note that this data appears to be anonymized, meaning that the "Unnamed" column likely does not correspond to a patient's name.

```

Confusion Matrix :
[[ 8  0]
 [ 0 75]]
Accuracy Score : 1.0
Report :
           precision    recall  f1-score   support

     0       1.00      1.00      1.00         8
     1       1.00      1.00      1.00        75

   accuracy                1.00         83
  macro avg              1.00      1.00      1.00         83
 weighted avg              1.00      1.00      1.00         83
    
```

This image shows a confusion matrix and classification report, which assess how well a machine learning model performs in classifying patients as having a disease or not (positive/negative) based on certain features.

The confusion matrix highlights the accuracy: all 80 positive cases and 75 negative cases were correctly classified, resulting in a perfect score of 1.0.



The classification report provides further details like precision, recall, and F1-score for each class, offering a more comprehensive picture of the model's performance.

In this case, the precision, recall, and F1-score are all 1.0 for both the positive and negative classes, which again indicates perfect classification performance.

Overall, the confusion matrix and classification report show that the machine learning model performed very well on this classification task. Ponders have appeared the show can accomplish great exactness, especially in foreseeing mortality chance. This permits specialists to prioritize high-risk patients for more forceful treatment. ([Important investigate can be found by looking for "time arrangement classification COVID-19 quiet results"]).

```

p = model.predict([X_test[80]])
p=round(abs(p[0]))
print('Predicted: %.3f' % p)
Predicted: 1.000

p = model.predict([X_test[20]])
p=round(abs(p[0]))
print('Predicted: %.3f' % p)
Predicted: 1.000

p = model.predict([X_test[66]])
p=round(abs(p[0]))
print('Predicted: %.3f' % p)
Predicted: 1.000

core = cross_val_score( model,X_train_res,y_train_res,cv = 5,scoring = 'accuracy')
print("The accuracy score of {} is:".format(model),score.mean())
The accuracy score of Lasso() is: 1.0

```

The image is a snippet of Python code that appears to be evaluating a machine learning model's performance. Let's break down the code line by line:

1. `p = model.predict([X_test[80]])`: This line predicts the class label for the 81st element (index 80) in the test data (`X_test`) using the trained model (`model`). The predicted class label is stored in the variable `p`.
2. `p = round(abs(p[0]))`: This line takes the absolute value of the first element in `p` (which likely contains the predicted probability for a class) and rounds it to the nearest integer. The result is stored back in `p`.
3. `print('Predicted: %.3f' % p)`: This line prints a message "Predicted:" followed by the value stored in `p`, formatted to three decimal places.

Lines 4-9 follow the same pattern, predicting the class label for different elements (20th and 66th) in the test data and printing the predicted values.

This line of code (`core = cross_val_score(model, X_train_res, y_train_res, cv = 5, scoring = 'accuracy')`) performs a technique called 5-fold cross-validation to evaluate the performance of the machine learning model (`model`) on the training data (`X_train_res` and `y_train_res`). Cross-validation is like a practice test for your machine learning model. It helps us understand how well the model would perform on completely new data it hasn't seen before. Here, the test focuses on accuracy, meaning the cross-validation process checks how many predictions the model gets right on different subsets of the training data.

`print("The accuracy score of {} is:".format(model), score.mean())`: This line prints the average accuracy score obtained from the cross-validation process.

Overall, this code snippet seems to be evaluating a machine learning model's performance on a classification task using cross-validation and printing the accuracy score.



## V. DISCUSSIONS

### Early Prediction Capabilities

The show can make expectations based on information from fair the primary 24 hours of hospitalization. This empowers altogether priorintercession compared to conventional strategies, possibly moving forward quiet results.

### Resource Optimization

By successfully classifying patients based on anticipated seriousness, the show can direct assignment of basic assets like ICU beds and ventilators to those in most basic require.

It's vital to consider that:

Assist inquire about is likely required to affirm and refine the model's adequacy in bigger and more different quiet populaces.

The particular execution measurements (precision, affectability, specificity) may not be promptly accessible in all ponders.

In general, the comes about recommend this time-series based approach holds guarantee for moving forward treatment methodologies and asset assignment in COVID-19 patients.

## VI. CONCLUSION

Machine learning offers promising tools to combat COVID-19. Techniques like lasso regression can analyze readily available clinical data to predict disease severity in patients with comorbidities. This can significantly aid healthcare professionals by enabling personalized medicine and improved resource allocation. While initial studies using datasets like the one described (273 patients with parameters like age and vitals) provide a starting point, further research is necessary.

Artificial neural networks present another avenue for analysis, but challenges like choosing informative data and ensuring generalizability to new patient populations require careful consideration. Overall, machine learning holds immense potential for improving COVID-19 prognosis and treatment, and future research should focus on addressing these challenges to optimize its effectiveness.

## REFERENCES

- [1] Adekunle Sanyaolu, Chuku Okorie, Aleksandra Marinkovic, Risha Patidar, Kokab Younis, Priyank Desai, Zaheeda Hosein, Inderbir Padda, Jasmine Mangat Risha Patidar, "Comorbidity and its Impact on Patients with COVID-19". Springer link publisher, Vol 2, 1069–1076 (2020).
- [2] Thakur, J S; Paika, Ronika; Singh, Sukriti<sup>1</sup>; Narain, Jai Prakash<sup>2,3</sup>, "COVID-19 and noncommunicable diseases: Impact and the strategic approaches", Wolters Kluwer, International Journal of Noncommunicable Diseases 5(2):p 29-35, Apr–Jun 2020.
- [3] Shweta Jakhmola,† Omkar Indari,† Budhadev Baral,† Dharmendra Kashyap, Nidhi Varshney, Ayan Das, Sayantani Chatterjee, and Hem Chandra Jha, "Comorbidity Assessment Is Essential During COVID-19 Treatment", frontiers in physiology Publisher, Volume 11 - 2020 .
- [4] Bolin Wang<sup>1</sup>, Ruobao Li<sup>2</sup>, Zhong Lu<sup>3</sup>, Yan Huang<sup>3</sup>, "Does comorbidity increase the risk of patients with COVID-19: evidence from meta-analysis", Aging Publishers, Volume 12, Issue 7 pp 6049–6057(2020)
- [5] Xiaoyu Fang<sup>1,2</sup>, Shen Li<sup>3</sup>, Hao Yu<sup>4</sup>, Penghao Wang<sup>3</sup>, Yao Zhang<sup>2</sup>, Zheng Chen<sup>2</sup>, Yang Li<sup>5</sup>, Liqing Cheng<sup>6</sup>, Wenbin Li<sup>7</sup>, Hong Jia<sup>1</sup>, Xiangyu Ma<sup>2</sup>, "Epidemiological, comorbidity factors with severity and prognosis of COVID-19: a systematic review and meta-analysis", Aging Publisher, Volume 12, Issue 13 pp 12493–12503 (2020).
- [6] Lin Wang, Zheng Yin, Mamta Puppala, Chika Ezeana, Kelvin Wong, Tiancheng He, Deepa Gotur, Stephen Wong, "A Time-Series Feature-Based Recursive Classification Model to Optimize Treatment Strategies for Improving Outcomes and Resource Allocations of COVID-19 Patients", IEEE publisher, vol. 26, no. 7, pp. 3323-3329, 2022.
- [7] Mohamed Abd Elaziz, Khalid M. Hosny, Ahmad Salah, Mohamed M. Darwish, Songfeng Lu, Ahmed T. Sahlol, "New machine learning method for image based diagnosis of COVID- 19", PLOS ONE publisher, Volume 15, Issue 6, 2020.



- [8] Wei-jie Guan, Ph.D., Zheng-yi Ni, M.D., Yu Hu, M.D., Wen-hua Liang, Ph.D., Chun-quan Ou, Ph.D., Jian-xing He, M.D., Lei Liu, M.D., Hong Shan, M.D., Chun-liang Lei, M.D., David S.C. Hui, M.D., Bin Du, M.D., Lan-juan Li, M.D., Guang Zeng, M.Sc., Kwok-Yung Yuen, Ph.D., Ru-chong Chen, M.D., Chun-li Tang, M.D., Tao Wang, M.D., Ping-yan Chen, M.D., Jie Xiang, "Clinical Characteristics of Coronavirus Disease 2019 in China", *The NEW ENGLAND JOURNAL of MEDICINE*, VOL. 382 NO. 18, 2020.
- [9] Joseph Friedman, Patrick Liu, Christopher E. Troeger, Austin Carter, Robert C. Reiner Jr, Ryan M. Barber, James Collins, Stephen S. Lim, David M. Pigott, Theo Vos, Simon I. Hay, Christopher J. L. Murray & Emmanuela Gakidou, "Predictive performance of international COVID-19 mortality forecasting models", *nature communications publisher*, Vol 12, pp.no 1-13, 2021.
- [10] Jianxi\_Luo, "Forecasting COVID-19 pandemic: Unknown unknowns and predictive monitoring", *ScienceDirect publisher*, Vol 166, 2021.
- [11] David Monterde, Gerard Carot-Sans, Miguel Cainzos-Achirica, Sònia Abilleira, Marc Coca, Emili Vela, Montse Clèries, Damià Valero-Bover, Josep Comin-Colet, Luis García-Eroles, Pol Pérez-Sust, Miquel Arrufat, Yolanda Lejardi & Jordi Piera-Jiménez, "Performance of Three Measures of Comorbidity in Predicting Critical COVID-19: A Retrospective Analysis of 4607 Hospitalized Patients", *Taylor & Francis Online*, Volume 14, 2021.
- [12] Mohitosh Biswas, Shawonur Rahaman, Tapash Kumar Biswas, Zahirul Haque, Baharudin Ibrahim, "Association of Sex, Age, and Comorbidities with Mortality in COVID-19 Patients: A Systematic Review and Meta-Analysis", *Karger*, Vol 64, Issue 1, pp.no 36-47, 2021.