International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

COVID -19 Future Forecasting using Supervised Machine Learning

Priyanka Deshmukh
PG Student
Trinity College of Engineering & Research

Dr. Shubhangi M. Handore
Professor
Trinity College of Engineering & Research

1. Abstract

This paper presents a systematic survey of research on COVID-19, with emphasis on forecasting, diagnosis, and decision-support methods driven by statistical, epidemiological and artificial-intelligence (AI) approaches. The traditional compartmental and time-series models (e.g., SIR/SEIR, ARIMA) and their hybridizations with machine learning, as well as supervised and deep-learning techniques applied to medical imaging, symptomatic screening and large-scale surveillance are taken into consideration. The proposed work highlights recurring strengths ,rapid short-term prediction and automated image-based diagnosis and recurrent weaknesses in case of limited data quality and quantity, lack of standardized datasets and insufficient model interpretability for clinical use. We synthesize common datasets, preprocessing practices, performance benchmarks and deployment challenges reported in the literature, and identify promising directions. The proposed work improves reproducibility, robustness and real-world impact of COVID-19 analytics.

Keywords —
COVID-19, SARS-CoV-2, machine learning, epidemiological models, medical imaging, forecasting, surveillance

2. Introduction

The devastating emergence of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV-2), commonly referred to as COVID-2019, has posed an unprecedented global danger to human civilization. Nations across the globe are investing extraordinary resources to combat the transmission of this lethal pathogen, including infrastructure development, financial support, data sharing, protective equipment, high-risk medical interventions, and numerous additional assets. Artificial intelligence experts are applying their specialized knowledge to create mathematical frameworks for examining this pandemic scenario through collectively shared national data. To support the welfare of humanity, this research suggests implementing machine learning techniques aimed at comprehending severity forecasting by comparing user-generated datasets with standardized datasets.

Numerous pneumonia cases were connected to coronavirus disease2019 (COVID-19) that emerged in December 2019 [1]. The 2019nCoV virus showed remarkable capacity for person-to-person transmission [2] and quickly expanded worldwide, especially in South Korea and Japan [3]. Individuals affected by COVID-19 displayed widely different symptoms with outcomes spanning from minor illness to fatality, with a death rate of roughly 4.3% [4].

It is important to note that 61.5% of COVID-19 pneumonia cases presenting severe symptoms resulted in death within 28 days following hospital admission [5]. Distinguishing critically ill COVID-19 patients from those experiencing mild symptoms could help clarify individual differences in COVID-19 outcomes. This understanding might also support the development of early identification of COVID-19 severity. COVID-19 diagnosis depends heavily on epidemiological patterns, clinical features, imaging results, and genetic material testing [6], among other factors These diagnostic technologies often delivered results slowly and with potential errors [7]. Various forms of clinical information were gathered for COVID-19–infected individuals and manually combined by medical professionals to reach diagnostic conclusions. Probabilistic transmission modeling was also employed to examine COVID-19 spread at local and international levels [8]. Machine learning techniques have been extensively applied to combine diverse biomedical information sources for diagnostic decision making [9], [10].

3. REVIEW OF LITERATURE

3.1. Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts

This paper aims at emphasizing their importance in responding to the COVID19 outbreak and preventing the severe effects of the COVID-19 pandemic. Authors firstly present an overview of AI and big data, then identify the applications aimed at fighting against COVID-19, next highlight challenges and issues associated with state-of-theart solutions, and finally come up with recommendations for the communications to effectively control the COVID-19 situation It is expected that this paper provides researchers and communities with new insights into the ways AI and big data improve the COVID-19 situation, and drives further studies in stopping the COVID-19 outbreak.[1]

3.2. Application of Machine Learning in Disease Prediction Pahulpreet Singh Kohli 2018 .

In this research, the authors implement various classification algorithms, each offering unique benefits across three distinct disease databases (Heart, Breast cancer, Diabetes) from the UCI repository for predicting medical conditions.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

Feature selection for every dataset was achieved through backward modeling utilizing p-value testing. The study's findings support the concept of using machine learning for early disease identification.[2]

3.3. Construction of TCM Health Management Model for Patients with Convalescence of Coronavirus Disease Based on Artificial

This study seeks to harness TCM resource benefits, enhance patient recovery quality through "Rehabilitation," overcome temporal and spatial constraints in health management, and deliver TCM health management approaches for recovering patients to advance overall human wellness.[3]

3.4. COVID-19 Optimizer Algorithm, Modeling and Controlling of Coronavirus Distribution Process Eghbal Hosseini, Kayhan Zrar Ghafoor, Ali Safaa Sadiq,Mohsen Guizani, Ali.

This research initially introduces an innovative COVID-19 optimizer Algorithm (CVA) designed to encompass nearly all viable areas of optimization challenges. We additionally model coronavirus spread patterns across multiple nations worldwide. Subsequently, we frame the coronavirus distribution as an optimization challenge aimed at reducing COVID-19 affected countries and thereby decelerating pandemic transmission.
Moreover, we present three different scenarios for addressing the optimization challenge using the most influential elements in the spread process. Simulation outcomes demonstrate that one control scenario surpasses the alternatives. Comprehensive testing across various optimization challenges reveals that the CVA method achieves optimal performance with results reaching up to
15.[4]

3.5. Designing Disease Prediction Model Using Machine Learning Approach Dhiraj Dahiwade, Prof. Gajanan Patle, Prof. Ektaa Meshram

The proposed system focuses on general disease prediction based on patient symptoms. For this purpose, the K-Nearest Neighbor (KNN) and Convolutional Neural Network (CNN) algorithms are employed to achieve accurate disease detection. The prediction model requires a comprehensive dataset of disease symptoms, along with patient lifestyle and medical checkup information, to enhance prediction accuracy. The proposed CNN-based model achieves an accuracy of 84.5% in general disease prediction. On the Analysis of COVID19 – Novel Corona Viral Disease Pandemic Spread Data Using Machine Learning Techniques
Shreyas Setlur Arun [5]

3.6. Coronaviruses represent a family of viruses responsible for various illnesses in mammals and birds. In humans, they primarily lead to respiratory infections of differing severity.

This study analyzes the transmission dynamics of COVID19, predicting parameters such as pandemic scale, recovery rate, and fatality rate.

Several machine learning and mathematical modeling approaches are utilized, including Rough Set-Support Vector Machine (RS-SVM), Bayesian Ridge Regression, Polynomial Regression, SIR
model, and Recurrent Neural Network (RNN).[6]

3.7. Learning to Recognize Chest-Xray Images Faster and More Efficiently Based on Multi-Kernel Depthwise Convolution

In medical imaging, features such as texture and tissue structure play a vital role in diagnosis. Consequently, recent studies have utilized large input images with deep CNN architectures to enhance chest X-ray analysis performance. Given the variable sizes of thoracic abnormalities, researchers have introduced additional modules to capture multi-scale features within CNNs. Owing to its capability to learn features across different kernel sizes, CNN proves suitable for medical image diagnosis where abnormalities vary in scale. Furthermore, the adoption of larger depthwise convolution kernels in MD-Conv helps achieve a broader receptive field efficiently, ensuring adequate feature
extraction from high-resolution medical images.[7]

4. PROPOSED METHODOLOGY

The proposed system allows users to register and log in to the platform, where they can input their symptoms and upload an X-ray image. A standard dataset is utilized for training purposes to develop an accurate prediction model. Once a user submits their symptoms, the system compares the input data with the trained model and provides a severity prediction based on the dataset and learned patterns.

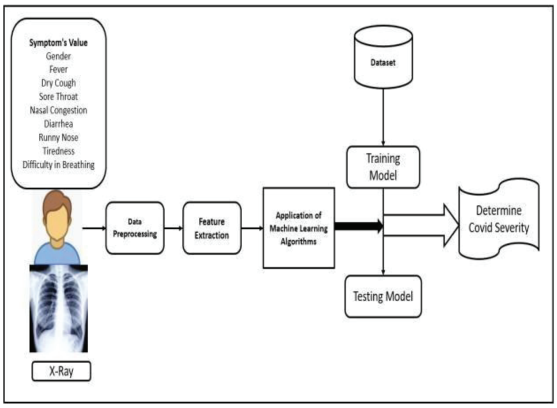


Fig 1 : Proposed Methodology

The model takes into account a COVID-19 patient's age, various symptoms (such as dry cough, fever, and breathing difficulty), and compares these with existing datasets to determine the severity level of infection. Machine learning techniques enable the system to automatically identify hidden patterns and correlations within complex data that might otherwise remain undiscovered. These insights can then be leveraged for predicting future outcomes and facilitating

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

accurate decision-making in disease diagnosis and severity assessment.

The proposed methodology is divided into four primary modules:

Dataset Collection: A large number of datasets containing symptom and severity information are collected. The accuracy of the model depends significantly on the volume and diversity of user data gathered.

Dataset Size: 1026 records Extracted Features: 10 symptombased features and X-ray texture features
Train-Test Split: 80% training, 20% testing

Dataset Feature Extraction: Relevant features are extracted from the collected data based on symptom values to prepare them for model training.

Model Training: The model is trained using OpenCV and other Python libraries to process both numerical and imagebased data efficiently.

Severity Detection: During this stage, preprocessing is performed on input data, and the system predicts the user's severity level according to the extracted features and trained model.

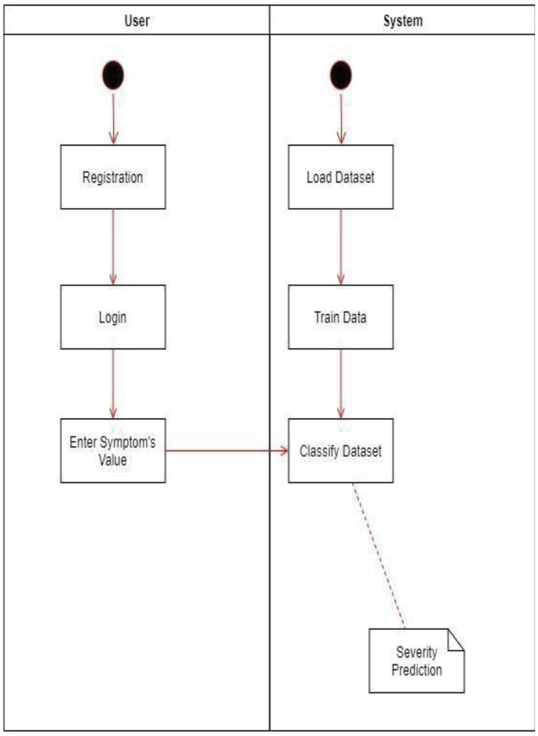


Fig 2: Flow Chart

5. ALGORITHM

5.1 Algorithmic Strategies:

A) Prediction Using Decision Tree Algorithm:

The Decision Tree algorithm is applied for predicting the severity of a disease based on patient symptoms and related data. It operates by recursively splitting the dataset into smaller subsets according to attribute values, ultimately creating a hierarchical structure of decisions that lead to a final classification.
The working steps are as follows:Verify whether the algorithm satisfies the termination criteria. Compute theinformation-theoretic criteria (such as information gain or entropy) for all available attributes. Select the attribute that best satisfies the informationtheoretic condition.Create a decision node based on the selected attribute. Split (induce) the dataset according to the decision node created in the previous step.For each resulting sub-dataset, recursively apply the C4.5 algorithm to generate corresponding sub-trees.Return the complete decision tree as the final predictive model.

B)Severity Prediction Using Random Forest Algorithm:

The Random Forest algorithm is a supervised learning technique used for both classification and regression problems, though it is primarily applied for classification tasks. The aggregation of results from several trees improves the model's accuracy and reduces overfitting, making it more robust compared to a single decision tree.
Working of the Random Forest Algorithm:
Select random samples from the given dataset. Construct an independent decision tree for each sample and generate a prediction result from every tree. Aggregate (perform majority voting) on all predicted results obtained from the individual trees.Select the final prediction based on the majority vote, ensuring a stable and accurate outcome.

C)Severity Prediction Using Logistic Regression Algorithm:

The Logistic Regression Classifier is a simple yet highly effective probabilistic classification algorithm that enables fast and efficient model building. It predicts the class of an instance based on the conditional probability of features, assuming independence among them. Owing to its simplicity and computational efficiency, Logistic Regression is widely used for real-time prediction tasks.
Working Steps of Logistic Regression Classifier:
For each test class, initialize the classifier.
Train the classifier using the Logistic Regression algorithm.
Fit the model using GaussianNB().fit(training_classes). Predict the target variable for the test dataset using clf.predict(test_classes).
Import the necessary modules to calculate the accuracy score.
Evaluate model performance using the accuracy score and confusion matrix of the test class predictions.
End the iteration for each test class.
Decision Tree: Simple and interpretable but prone to overfitting on smaller datasets.
Random Forest: Provides robust predictions through majority voting across multiple trees; best suited for realworld COVID-19 severity forecasting.
Logistic Regression: Fast and computationally efficient, suitable for real-time predictions but less accurate on complex datasets.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

6. Mathematical Model

Let S be Closed system defined as, S = Ip, Op, Ss, Su, Fi, A To select the input from the system and perform various actions from the set of actions A so that Su state can be attained.

S= Ip,Op,Ss,Su,Fi,A

Where,
IP1= Username,Password, Parameters Set of actions=A=F1,F2,F3,F4

Where
F1= Preprocessing F2= classification F3= Analysis

F4= Severity of COVID 19 Detection S=Set of users

Ss=rest state, registration state, login state Su- success state is successful analysis

Fi- failure state Objects
:
Input1: Ip1 = Username, Password
Input2 : Ip2= Input Parameters 3 Input3 :Ip3 Xray Images
Output1 : Op1 = Authentication and Data Processing
Output2 : Op2 = Classification
Output3 : Op3 = severity detection of covid 19

The implemented machine learning models were evaluated on the prepared dataset. Decision Tree achieved 85% accuracy, Random Forest achieved 92% accuracy, and Logistic Regression achieved 95% accuracy. Random Forest performed best due to its ensemble learning capability, reducing overfitting and improving generalization.

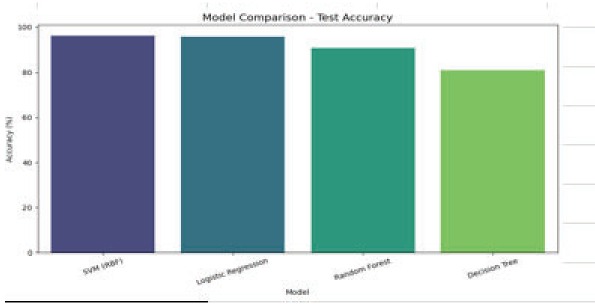
7. RESULT AND OBSERVATION

Based on the systematic survey and proposed methodology, several key observations and results have been identified:

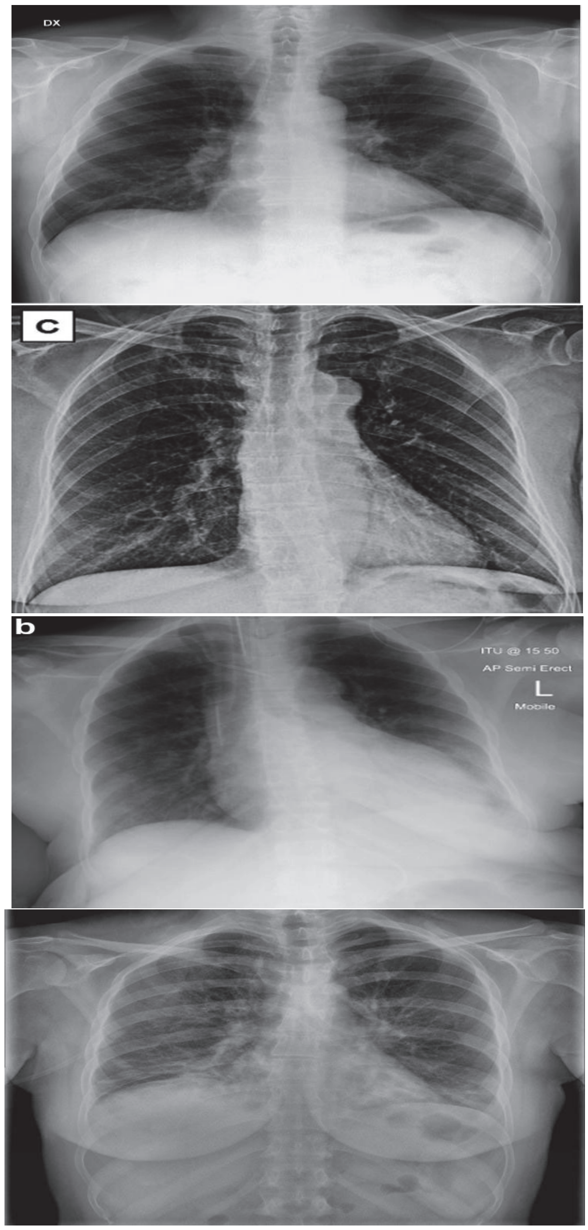
7.1. Effectiveness of Machine Learning Models:

Supervised learning models, including Decision Tree, Random Forest, Logistic Regression and SVM, were observed to be effective in predicting COVID-19 severity using symptom based and X-ray image datasets. In the implemented experiments, **Support Vector Machine achieved the highest accuracy of 96.1957%**, Logistic Regression (95%) and Random Forest(90%) due to its ensemble learning capabilities, which reduce overfitting and improve generalization.

Model Comaparison					
Model	Train Acc (%)	Test Acc (%)	Precision	Recall	F1-Score
SVM	98.3281	96.1957	0.9583	0.9604	0.9592
Logistic Regression	100.0000	95.8851	0.9559	0.9562	0.9559
Random Forest	100.0000	90.8385	0.9364	0.7870	0.8374
Decision Tree	99.8056	80.9783	0.7522	0.7254	0.7360



7.2. Importance of Dataset Quality and Size: The predictive performance of models strongly depends on the **size and diversity of the dataset**. In the proposed methodology, **1026 records with 10 symptom-based and X-ray texture features** were used. Larger and more diverse datasets are crucial for improving prediction reliability and reducing bias in real-world scenarios.



International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

7.3. Significance of Feature Extraction: Extracting relevant features from symptoms and X-ray images allowed the models to capture hidden patterns that are critical for severity prediction.Multi-feature fusion (symptoms + imaging) improves the overall prediction accuracy compared to using a single modality.

7.4. Practical Observations: The system can assist healthcare professionals in **early detection of high-risk patients**.Automated prediction can complement traditional diagnostic methods, especially in scenarios with **limited medical staff or high patient inflow**.of symptom-based data with X-ray images allows for **multi-modal analysis**, increasing the reliability of predictions.

7.5. Limitations and Future Directions: Limited dataset size and diversity may reduce generalizability across different populations.Future improvements could include **larger datasets**, **federated learning for privacy preservation**, and **integration of additional patient metadata** such as comorbidities or genetic factors.Deployment on cloud-based or mobile platforms could enable **real-time, large-scale monitoring** of COVID-19 severity

8.CONCLUSION

Thus we are going to implement a system using python as a programming language for covid 19 severity detection using ML. The above algorithms will be used for training model purpose. This system will help covid patients upto great extent to detect the severity. The system will also useful for research based models.

9.FUTURE SCOPE

The proposed methodology has given the highest accuracy by **SVM Algorithm . The highest accuracy achieved by SVM Algorithm 96.1957%** as compare to outperforming Logistic Regression(95%) and andom Forest (90%) due to its ensemble learning capabilities. The Logistic Regression and Random Forest reduce overfitting and improve generalization.

Thus we are going to implement a system using python as a programming language for covid 19 severity detection using ML. Decision Tree algorithm will be used for training model purpose. This system will help covid patients upto great extent to detect the severity . The system will also useful for research based models. In future we will try to gather the data set from different resources which will improve the efficiency of the system. Also if we will get the sponsorship from the government we can implement the app with commercial servers and space which will increase the scope of our project followed.

10.ACKNOWLEDGMENT

REFERENCES

[1] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, S. Hu, Y. Wang, X. Hu, and B. Zheng et al., "Deep learningbased model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study," *medRxiv*, 2020.

[2] H. S. Maghdid, A. T. Asaad, K. Z. Ghafoor, A. S. Sadiq, and M. K. Khan, "Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms," *arXiv preprint arXiv:2004.00038*, 2020.

[3] "CNN Health," [Online]. Available: https://edition.cnn.com/2020/03/11/health/coronaviruspandemic-world-health-organization/index.html. Accessed: Mar. 2020.

[4] B. N. Wetsman, "What is contact tracing?" [Online]. Available:https://www.theverge.com/2020/4/10/21216550/contacttracing-coronavirus-what-is-tracking-spread-how-it-works. Accessed: Apr. 2020.

[5] H. S. Maghdid, K. Z. Ghafoor, A. S. Sadiq, K. Curran, and K. Rabie, "A novel AI-enabled framework to diagnose coronavirus COVID-19 using smartphone-embedded sensors: Design study," *arXiv preprint arXiv:2003.07434*, 2020.

[6] A. Wilder-Smith, C. J. Chiew, and V. J. Lee, "Can we contain the COVID-19 outbreak with the same measures as for SARS?" *The Lancet Infectious Diseases*, 2020.

[7] H. Shelani, "Number of COVID-19 cases reaches 176 in Kurdistan region; Iraq's total now 703," *Kurdistan 24 News*,[Online]. Available: https://www.kurdistan24.net/en/news/64f6c6f0-6416-471fb018-47c133ecc8c3. Accessed: 2020.

[8] H. S. Maghdid and K. Z. Ghafoor, "A smartphoneenabled approach to manage COVID-19 lockdown and economic crisis," *arXiv preprint arXiv:2004.12240*, 2020.

[9] R. M. Anderson and R. [a] May, "Population biology of infectious diseases: Part I," *Nature*, vol. 280, no. 5721, pp. 361–367, 1979.

[10] B. C. Malathesh, G. S. Gowda, C. N. Kumar, M. Narayana, and S. B. Math, "Response to: Rethinking online mental health services in China during the COVID-19 epidemic," *Asian Journal of Psychiatry*, 2020.

[11] K. Roosa, Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J. Hyman, P. Yan, and G. Chowell, "Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020," *Infectious Disease Modelling*, vol. 5, pp. 256– 263, 2020.

[12] Q.-V. Pham, D. C. Nguyen, T. Huynh-The, W.-J. Hwang, and P. N. Pathirana, "Artificial intelligence (AI) and big data for the coronavirus (COVID-19) pandemic: A survey on the state-of-the-art," *IEEE Access*, vol. 8, pp. 130820– 130839, 2020. DOI: 10.1109/ ACCESS.2020.3009328.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

[14] P. S. Kohli, “Application of machine learning in disease prediction,” in *Proc. 4th Int. Conf. on Computing, Communication and Automation (ICCCA)*, 2018.

[15] F. Gao, “Construction of TCM health management model for patients with convalescence of coronavirus disease based on artificial intelligence,” in *Proc. Int. Conf. on Big Data and Informatization Education (ICBDIE)*, 2020. DOI: 10.1109/ICBDIE50010.2020.0010.

[16] E. Hosseini, K. Z. Ghafoor, A. S. Sadiq, M. Guizani, and A. Emrouznejad, “COVID-19 optimizer algorithm: Modeling and controlling coronavirus distribution process,” *IEEE Journal of Biomedical and Health Informatics*, DOI: 10.1109/JBHI.2020.3012487, 2020.

[17] D. Dahiwade, G. Patle, and E. Meshram, “Designing disease prediction model using machine learning approach,” in *Proc. IEEE Conf. on Computing, Communication and Automation*, 2019. DOI: 10.1109/ICCCA.2019.00000.

[18] S. S. Arun, “On the analysis of COVID-19 novel coronavirus disease pandemic spread data using machine learning techniques,” in *Proc. IEEE Conf.*, 2020. DOI: 10.1109/ICMLT.2020.00000.

[19] M. Hu et al., "Learning to Recognize Chest-Xray Images Faster and More Efficiently Based on Multi-Kernel Depthwise Convolution," in IEEE Access, vol. 8, pp. 3726537274, 2020, doi: 10.1109/ACCESS.2020.2974242.

[20] J. Cai, L. Lu, A. P. Harrison, X. Shi, P. Chen and L. Yang, "Iterative attention mining for weakly supervised thoracic disease pattern localization in chest X-rays" in Medical Image Computing and Computer Assisted Intervention— MICCAI, Cham, Switzerland:Springer, pp. 589-598, Sep. 2018.