JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 2267-2280
https://publishoa.com
ISSN: 1309-3452

Review Of Machine Learning-Based Disease Diagnosis and Severity Estimation of Covid-19

Mr. Shaik Khasim Saheb
Research Scholoar, Dept.of CSE,
Annamalai University, Chidambaram, Tamilnadu, India
Assist. Professor, Department of CSE,
Sreenidhi Institute of Science and Technology, Yamnampet, Hyderabad,India
shaikkhasims@sreenidhi.edu.in

Dr. B. Narayanan
Assistant Professor, Computer Science and Engineering Wing,
D.D.E, Annamalai Univeristy, Chidambaram,Tamilnadu, India
narayanan.bk@gmail.com

Dr. Thota Venkat Narayana Rao
Professor, Department of Computer Science and Engineering
HoD-CSE IOT, Sreenidhi Institute of Science and Technology, Yamnampet, Hyderabad, India
venkatnarayanaraot@sreenidhi.edu.in

ABSTRACT
Following the identification of SARS-CoV-2, the novel coronavirus responsible for COVID-19, health care professionals have been pushed to develop novel technical solutions and patient treatment techniques. The COVID-19 outbreak has accelerated the deployment of machine learning (ML) technology, and various groups have been eager to embrace and adjust these ML solutions to the pandemic's concerns. We conducted a thorough assessment of the available literature on the use of machine learning in the fight against COVID-19, focusing on illness development, diagnosis, severity estimation, drug and treatment analysis, novel feature selection, and the post-Covid context. A systematic search of online research repositories such as Google Scholar, PubMed, and Web of Science was undertaken in accordance with the "Preferred Reporting Items for Systematic Reviews and Meta-Analysis" criteria to identify all relevant published papers between 2020 and 2022. The search syntax was created by combining COVID-19-specific terms with the word "machine learning."

Keywords: Disease Diagnosis, Machine learning, World Health Organization, Support Vector Machine, Deep Forest, Coronavirus Disease 2019, Chest X-rays, Extreme Gradient Boosting.

1. Introduction

COVID-19 is a viral pneumonia caused by the coronavirus 2 [1], which causes serious acute respiratory disease. The disease was initially discovered in 12th month of 2019 in Wuhan, China. Since then it has spread worldwide [2]. Nearly 22 million people have been infected as of August 19, with 773,067 fatalities reported in more than 200 countries [3]. The outbreak has been classified as a Worldwide Public Health Emergency and Epidemic by the World Health Organization (WHO) [4.]. Forecasting the epidemic's long-term pattern can help health officials figure out how the virus spreads and devise effective preventative and containment tactics. To forecast the progression of COVID-19, several researchers used both classic epidemiological models and machine learning technologies [5].

In their current state, the COVID-19 screening tests are not without defects and may occasionally provide incorrect findings. For a variety of pandemic situations, enhanced capability for screening huge populations as well as deploying containment strategies is required. Forecasting analytics for the likelihood of COVID

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 2267-2280
https://publishoa.com
ISSN: 1309-3452

screening positive could help with resource distribution and tracking down possible contacts [7]. Defining a machine learning technique that just needs basic data can aid in the testing of subgroups.

The most often utilized imaging modalities for detecting COVID-19 infections are CT, LUS, and chest X-ray imaging [8]. The visuals produced by clinical pneumonia and those produced by other viral infections, such as the influenza virus, have many similarities (Influenza A). The similarity of these situations makes clinical diagnosis difficult. Machine learning (ML)-based algorithms have shown to be extremely effective in the accurate analysis of medical images. In health-care settings. Machine learning (ML) methods are simple to automate and implement.

2. Disease Diagnosis and Severity Estimation using Machine Learning

The current literature on ML and AI-based covid-19 illness prediction and severity estimation has been comprehensively evaluated in this area. The common goal of these contributions is to forecast Covid-19 disease with low false alarms and human expert involvement, and machine learning severity estimation helps to reduce mortality by recognizing the requirement for in-hospital patient treatment to the required covid-19 infected patients.

Awal, M. A., et al., [11], have identified a timely answer for this epidemic. In the framework under examination, the COVID positive and negative classes of the dataset are equalized using Bayesian optimization. Despite the fact that the proposed method was only evaluated on nine current classifiers, it can be used to a wide range of classifiers and classification challenges. According to this study, XGBoost (Extreme-Gradient Boosting) has the maximum Kappa value (97 percent). When ADASYN is utilized, the kappa index improves by 96.94 percent. Bayesian optimization has also been likened to grid search and random search. SHAP analysis revealed that the data was dominated by a few key criteria. Other works have been cited in comparison. The proposed method might track COVID patients more quickly than current methods. Two potential applications have been demonstrated to assist clinicians and construct a recommendation system.

Machine learning was used to predict COVID-19 diagnosis and severity by de FátimaCobre et al., [12]. The COVID-19 test results of the patients were examined. Patients were classified based on the severity of their COVID-19 sickness. Outliers, trends, and significant variables in the data were investigated using exploratory analysis (PCA). Based on the clinical outcomes, machine learning algorithms were used to predict Covid positive and its sternness. This research used a variety of statistical models. Accuracy testing was performed on all four models. Samples were gathered from 5,643 patients, with 557 of them being positive and the rest being categorized aas negative. Subset 2 included an additional 257 good outcomes. Negative samples were properly identified by models such as DT, ANN, KNN, and PLS-DA. 86 percent of the time By discriminating between patients with acute and non-severe sickness, accuracy was reached.

Lorenzen, S. S., et al., [13] looked at using machine learning to predict future critical care demands. An analysis of 42,526 Danish patients affectedby Covid-19 in the past. After n days, Random Forest (RF) models could predict ICU admission and mechanical ventilation risk. We did additional in-depth analysis for n of 5 and n of 10. For forecasting the chance of an inpatient stay in the critical care unit and the likelihood of mechanical ventilation, the models' ROC-AUCs ranged from 0.981 to 0.995. The capacity and ventilation of the ICU were anticipated to range from 0.333 to 0.9889 on average by the n-day forecasting models. More than a day in advance, forecasting models have failed to produce accurate forecasts (for large n). ROC-AUC of 0.990 and R2 of 0.928 predicted ICU capacity and ventilator usage, respectively, and ROC-AUC of 0.99 and R2 of 0.928 predicted ICU capacity and ventilator use.

Kar, S., et al., [14] constructed and validated individualised mortality risk scores based on data evaluated using patient-specific, non-identifying clinical data gathered upon admission to the hospital. Using an electronic medical record, they collected data from 1393 hospitalized patients (Expired—8.54%). The clinical and laboratory data of the patient were first obtained and analyzed. XGB was used to build the ML model (Extreme Gradient Boosting). For the Cox Regression Hazard Model, the XGB Algorithm was employed. There are 977 people in the prospective verification cohort (8.3 percent mortality). Its AUC ROC Score was 0.8685 and its Accuracy Score was 96.89, respectively.

M. A. Dabbah et al. [15] built a data-driven random forest classification model that performed very well using baseline features, pre-existing illnesses, symptoms, as well as vital signs (AUC: 0.91). According to the authors, COVID-19 mortality forecasting were as good as or better than pre-existing high-risk chronic conditions like comprehensive anthropometric measurements and acute renal failure, UTI (urinary tract infection), as well as pneumonia.

Nakamichi, K., et al., [16] discovered that SARS-CoV-2 sequence variations were linked to COVID-19 patient outcomes in a single medical system. Researchers had to look no further than clinical materials to find the SARS-CoV-2 RNA genome sequence. The demographic and clinical information of patients was gathered, including when they were admitted to the hospital and when they died. Statistical and machine learning techniques were used to investigate the link between viral genetic variations and hospitalization or mortality. Twelve polymorphisms in five genes demonstrated that two major viral clades are unique from one another. Clade 2 infections were associated with a higher likelihood of hospitalization ($p = 0.06$). AUC of 0.93 was predicted by machine learning techniquestrained on demographics as well as co-morbidities of the individuals.

The ability of a machine learning method to predict invasive mechanical ventilation in COVID-19 patients within 24 hours was investigated by Burdick, H., et al. [17]. In total, 197 people were enrolled in the intended study. According to Burdick, H., the algorithm had a better diagnostic odds ratio (12.58) than the Modified Early Warning Score for predicting ventilation (MEWS). Furthermore, the algorithm's sensitivity (0.90) was higher than MEWS' (0.78), as was its specificity (p 0.05).

An ML-based tool was built using EHR data to maximize clinical utility while staying within a constrained budget for the purpose of predicting bad outcomes, according to Nguyen, S., et al., [18]. They were able to obtain a deeper grasp of how machine learning models generate judgments. This study employed de-identified COVID-19 EHR data. The Nguyen, S. examined multiple machine learning approaches for predicting increased ventilator support or death. According to Nguyen, S., predictive performance may decrease at the expense of large cost savings.

An explainable machine learning system, according to Casiraghi, E., et al., [19], could give simple decision criteria to aid doctors in determining patient risk. Boruta and Random Forest (RF) are used in ten-fold cross-validation methods to create variable significance ratings. In the final stage before building an associative tree, RF classifier rules are retrieved, simplified, and pruned. The findings demonstrate that the number of comorbidities as well as laboratory factors can aid in risk prediction.

Boussen, S., et al., [20] established a method to predict COVID-19 patient length of stay using breathing frequency (BF) and oxygen saturation (SpO_2) measurements. In COVID-19 patients admitted to the ICU, they monitored BF and SpO_2. The data was clustered using Gaussian mixture. The algorithm was put to the test to see if it could resist real-world intubation rates. The proposed approach detected need of ICU admission with an accuracy of 87.8 percent (TPR = 86.5 percent, TNR = 90.9 percent).

Ponce, D., et al., [21] looked at the following categories to see whether there were any prediction characteristics for AKI: AKI causes and symptoms. After fitting models to the training data, they measured the AUC-ROC to determine accuracy. Elastic Net coefficients were utilized to assess the Covid-19 scope in AKI patients. In the validation cohort, the AUC-ROC was 0.823.

Rechtman, E., et al., [22] looked at a total of 8770 SARS-CoV-2 patients. Conventional machine learning based regression techniques were used to analyse demographic, clinical, and comorbidity factors. Male sex, older age, higher BMI, increased heart rate and respiratory rate, and drowned oxygen saturation were all connected to COVID-19 mortality. These factors shown to be potential to predict the mortality caused by COVID-19.

The participants in the Xu, W., et al. [23] study were 659 COVID-19 patients. ML approaches were used to create prediction models based on the medical parameters of ARDS and non-ARDS. Artificial intelligence techniques based on conventional and deep learning. ARDS patients were 7.5 years older than non-ARDS patients, with a median age of 56.5 years. The best AUC, sensitivity, specificity, and accuracy were achieved by modelling and effect evaluation using a range of AI techniques to identify mild persons at risk of developing ARDS.

To forecast the need of ICU admission, need of mechanical ventilation, and mortality in Covid-19 patients, Estiri et al. [24] used a mining model. It predicts future results based on the health records of prior patients. Over 600 factors from individuals' health records gathered before Covid-19, as well as demographics, were used to estimate the four undesirable outcomes. The AUC ROC for predicting death was 0.91, whereas the ROC for envisagingneed of ICU, and need of mechanical ventilation was 80% to 81%.

S. Roy et al. [25] used an international database for monitoring and reporting COVID-19 outcomes to develop clinical COVID-19 results for IBD patients. Machine learning was used to predict COVID-19 outcomes in IBD patients using primary and secondary factors. Factors like age, comorbidities, and medication use are thought to predict 70 percent of COVID-19/IBD results.

X-ray scans can be utilized to anticipate the severity of COVID-19 patient hazards, according to Sayed, S. A. F., et al. [26]. The suggested model was created using CheXNet deep pre-trained and hybrid handcrafted methods. Using a mix of PCA and RFE, features were prioritized. The top classifiers in the tests were PCA and RFE selected features (PCA + RFE). With a combined (PCA + RFE) feature set, the XGBoost classifier obtained 97 percent accuracy, 98 percent precision, 95 percent recall, and 100 percent ROC-auc. SVM obtained 97 percent accuracy, 96 percent precision, 96 percent recall, 96 percent f1-score, and 96 percent ROC-auc in terms of accuracy. 99.6% of pre-trained CheXNet features were properly detected using Extra Tree and SVM with RFE.

Duckworth, C., et al., [27] demonstrated explainable machine learning in the COVID-19 epidemic. Using pre-COVID-19 data, a ML techniques was developed to predict high-risk individuals who should be admitted to the emergency department. On pre-pandemic attendances and during the COVID-19 pandemic, the model performed well (AUROC 0.856). (AUROC 0.8559). A 95 percent confidence interval was discovered (AUROC of 0.826).

Heldt, F. S., et al., [28] looked at 879 SARS-CoV-2 positives. EHRs were used to collect totally anonymised demographic, physiological, and laboratory data. In-hospital mortality, acute-care admission, and the need for invasive mechanical ventilation were predicted using data from initial ED visits. It was used to keep track of patients' progress. 15 percent of patients required intensive care, 7% required mechanical ventilation, and 31% died while in the hospital. With AUC-ROC scores of 0.76–0.87 (F1 scores of 0.42–0.60), the Heldt, F. S., claim that their models learned from early clinical data.

Simple characteristics were utilized by Murri, et al., [29] to predict death in SARS-CoV-2 patients (elevated blood-count, C-reactive protein, BUN, increased sodium levels, as well as a decrease in SpO2). According to the study, accuracy and discrimination power were determined to be excellent.

The Du, R., et al., [30] employed ML to detect Covid-19 infection. They used a retrospective dataset of 5,148 patients from 24 Hong Kong hospitals to build an ML model that can classify COVID-19 as well ascommon pneumonia aetiologies. They put the model through its paces on three distinct outbreaks. In all three validation sets, the ML model correctly identified SARS-CoV-2 infection with high AUCs and specificity but low sensitivity.

D. Patel et al. [31] used demographical, clinical, as well as blood panel profiling data to develop ML technologies for predicting the need for critical care and ventilators. The AUC for ICU need was 0.80, and for mechanical ventilation need was 0.82, according to Patel, D. They claim that all three data types are crucial. According to Patel, D., the AUC was reduced by 0.12 units when the blood panel profile data was removed. Only five traits were used to train the predictors, and they performed just as well as those who were trained on all of them.

AlJame, M., et al., [32] were able to appropriately identify COVID-19 using clinical and regular laboratory data. Using a deep forest (DF) ensemble-based technique, several classifiers were deployed in different levels to increase the model's performance. We employ XGBoost, LightGBM, and Extra Trees in the cascade level. Two benchmark datasets were used to train and test the model. The proposed DF model achieves 99.5 percent accuracy, 95.28 percent sensitivity, and 98.96 percent specificity.

M. Kukar et al. [33] constructed a machine learning method for COVID-19 detection using routine blood tests from 5333 individuals with different pneumonia related viral and bacterial illnesses and 160 individuals with COVID-19-positive. After cross-validation, the AUC was 0.97. Conventional blood tests were utilized to detect COVID-19. The t-SNE visualisation demonstrated that blood parameters in individuals with a high COVID-19 course resembled those of a bacterial infection rather than the infection caused by viral.

J. Qu et al., [34], evaluated conventional blood tests from 300 patients using logistic regression model to measure and compare the predictive value of numerous prognostic factors for the development of acute COVID-19. The haemoglobin, lymphocyte count, as well as ferritin levels were the most sensitive indications of acute COVID-19 illness.

According to Zhang, J., et al., [35], patient demographics, comorbidities, and common lab findings are sufficient for effectively predicting individual COVID-19 positive diagnoses. When it came to determining the value of characteristics, the most relevant ones were found to be AST and oxygen saturation. In addition, for sub-population stratification, a single decision tree model was built.

C. Gangloff et al. [36] used routine clinical and laboratory data to develop and validate machine learning model using NN (neural networks), RF (random forest),, and LR (logistic regression) for COVID-19 diagnosis in hospital admissions. The model was built using a unique set of variables. AUC was used to determine the performance of the model. A total of 536 individuals were examined, with 106 diagnosed with COVID and the rest 430 are healthy. The AUC of RT-PCR as well as chest-CT increased from 85% to 93% and 77% to 89% respectively.

Hussain, L., et al. [37] sought to develop a radiomic features based artificial intelligence tool to classify COVID-19 patients using portable chest x-rays. The texture as well as morphological characteristics were examined. This was done using diversified ML based AI algorithms. Two- as well as multi-class categories were examined. The performance of classification models was evaluated using ROC-curve.

Schmidt, M., et al. [38] sought to create predictive survival models for COVID-19 ICU patients after their survival of 1 to 2 weeks. Using machine learning, the researchers constructed dynamic, therapeutically valuable models that may predict 90-day mortality. The three models were chosen, which have been trained on 15 ICU-entry factors from the D1, D7, and D14 data. While external validation of the SOSIC-score has still been required, according to the experiments done by authors demonstrated that their approach is significant to assist the treatment decisions and notifying family members to the expected outcome.

COVIDPEN, a transfer learning strategy, has been suggested by Jaiswal, A. K., et al. [39]. It is a trimmed EfficientNet model for the diagnosis of COVID-19 cases. The proposed model is further approximated by utilizing post-hoc analysis, which allows to describe the predictions. To demonstrate the efficacy of the proposed approach, chest radiographs as well as computed tomography (CT) scans were used in training and testing phases. According to test findings with multiple benchmark comparisons, the procedure is comparable and provides clinically comprehensible examples intended for healthcare practitioners.

Jha, S. K., et al. [40] sought to develop a machine learning strategy for disease prediction that incorporated effective feature generation, selection, and classification techniques. The authors stated that this technique has outperformed the existing approaches in diagnosing and forecasting each of the specified diseases. The proposed method had the highest recognition accuracy, scoring 99.12 percent for primary tumor recognition, 96.45 percent for breast cancer recognition, 94.44 percent for cryotherapy recognition, and 93.84 percent for audiology recognition, among other evaluation metrics. Additionally, it takes care of the dataset's missing values.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 2267-2280
https://publishoa.com
ISSN: 1309-3452

Using clinical information from a multitude of sources, including Kaggle, Gomathi, S., et al. [41] proposed a machine learning-based predictive method that predicts the survival of severe patients. The authors sought to conduct a comprehensive evaluation of COVID-19's diagnostic potential. Prior to targeted intervention and diagnosis, this work upgrades the cost-effective and rapid categorization and prediction of criticality and survival.CoxPH (cox proportional hazard) makes machine learning accessible to non-ML specialists, enhances ML efficiency, and accelerates machine learning research. COVID-19 prediction will improve future vaccine development by allowing researchers to track incidences, particularly in India. Because of the exponential expansion of machine learning implementations, a market for off-the-shelf machine learning solutions (DT, NB, KNN, LR) that can be utilized rapidly and without any prior knowledge has emerged. The concept of AutoML is introduced and applied to data analysis and the selection of the optimal illness prediction algorithm in this study.

He, F., et al. [42] constructed and validated machine learning methods using EHR () of the multiple patients across the healthcenters situated in different geographical locatiions. COVID-19 is a de-identified database of longitudinal electronic health records. Based on clinical parameters, the methods used to determine mortality of in-patients within the range of 28-days. The methods also equipped to forecast the need of ICU admission, need of mechanical ventilators and respiratory failure in the inpatients.

Monaghan, C. K., et al. [43] developed a machine learning (ML) model that predicts the likelihood of a previously unaware covid-19 infection being discovered in a haemodialysis (HD) patient after three days or more. The authors used patient records from a national network of dialysis health centres to develop a machine learning system that includes 81 variables to assess the likelihood of an adult HD patient contracting a previously unknown covid-19 infection. The greatest predictors were intradialytic weight gain, necessarily imply pre-HD temperature of the body, and shift in post-HD pulse rate. To prevent false positives, authors set the threshold for classifying observational data as infected or healthy at 0.80.

Blair, P. W., et al. discovered patterns of inflammation associated with COVID-19 in people with no communicable
diseases using network machine learning [44]. (NCDs).
The researchers have chosen positive participants whose samples were available between 15 to 28 days of symptom onset. The concentrations of 15 inflammatory protein biomarkers in the blood were determined using a broad dynamic range immunoassay. The Fisher's exact test was used to identify NCDs with a stronger correlation (0.05 significant level) all over clusters, and the Kruskal-Wallis test was used to compare the NCD frequency from each cluster to the NCD frequency of all other clusters.

S.No	**References**	**Model**	**classifier/Classsification**	**Dataset**	**Features**	**Accuracy %**
1	Awal, M. A., et al., [11]	machine learning-based framework	ADAptiveSYNthetic algorithm with Bayesian optimization	Kappa index, HIPAA	SHAP analysis was used to identify features.	97.17%
2	de FátimaCobre, A., et al., [12]	principal component analysis, PCA	ANN, DT, PLS-DA, and KNN	public Kaggle platform	patients' laboratory tests results	84%
3	Lorenzen, S. S., et al., [13]	COVID-19 pandemic in Denmark	Random Forest	SARS-CoV-2 PCR	binary features	95%
4	Kar, S., et al., [14]	determine the likelihood of mortality	XGB	Hosmer and Lemeshow Goodness of Fit (GOF	The Cox Proportional Hazard Model in conjunction	96.89%

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 2267-2280
https://publishoa.com
ISSN: 1309-3452

					with the XGB Algorithm	
5	Dabbah, M. A., et al., [15]	a UK Biobank study on COVID 19 mortality	random forest classification model	UK Biobank	top-seven RF-ranked features	0.91%
6	Nakamichi, K., et al., [16]	A link between hospitalization and death	Random Forest	University of Washington Medicine health system	Demographics and clinical information	93%
7	Burdick, H., [17]	Prediction of respiratory decompensatio n	XGBoost	Pytho	flow features	86%
8	Nguyen, S., et al., [18]	Detection of negative effects in COVID-19 patients as early as possible.	LR, XGBoost, and Gaussian p	In Northwest Ohio and Southeast Michigan, ProMedica Health System	Entries of EHR	72% to 80%
9	Casiraghi, E., et al., [19]	Emergency Departments' COVID-19 Risk Prediction	Boruta and RF	Johns Hopkins University)	radiological features	76% to 81%
10	Boussen, S., et al., [20]	unsupervised machine learning technique	Gaussian mixture model (GMM)	teaching hospital, France	Spirometry (SpO2) and breathing frequency (BF).	87.8%
11	Ponce, D., et al., [21]	Acute renal damage patients in the COVID-19 mortality	XGBoost RF, Elasticnet	Latin America AKI COVID-19 Registry	demographic data, laboratory reports, AKI characteristics	95%
12	Rechtman, E., et al., [22]	Initial clinical encounters can predict COVID-19 mortality based on vital signs.	gradient-boosting	Mount Sinai Health System facilities NYC	Demographic and clinical characteristic, CKD, and COPD features	86%
13	Xu, W., et al., [23]	COVID-19 patient ARDS risk factor analysis	DT, RF, and SVM	Data on 659 individuals from 11 areas of China with COVID-19	Clinical characteristic, radiomic features	97 to 99%

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 2267-2280
https://publishoa.com
ISSN: 1309-3452

14	Estiri, H., et al., [24]	An iterative approach of selecting features and machine learning algorithms to predict health outcomes	MLHO	Nearly 13,000 patients have tested positive for COVID-19, making up a huge cohort.	clinical and demographic data	80% to.81%
15	Roy, S., et al., [25]	IBD patients have a higher COVID-19 death rate.	RF, LR	COVID-19/IBD repository	IBD related demographic characteristics	70%
16	Sayed, S. A. F., et al., [26]	CheXNet deep pre-trained model	XGBoost, Extra Tree, and SVM with RFE classifiers	X-ray image dataset, Bootstrapped	Radiomic features	98%
17	Duckworth, C., et al., [27]	Determines wh ich patients require hospitalization.	explainable machine learning (SHAP)	UHS	relation between the features of episodes and their related clinical effects	95%
18	Heldt, F. S., et al., [28]	Risk evaluation for COVID-19 patients at an early stage	Multivariate LR, RF, XGBoost.	NHS hospital trust, longitudinal information	64 clinical features	76% to 87%
19	Murri, R., et al., [29]	machine-learning risk prediction model	SHapley Additive exPlanations (SHAP	COVID-19 positive patients admitted to Fondazione PoliclinicoGemel li	Demographic and clinical characteristics	86%
20	Du, R., et al., [30]	covid-19 infection diagnosis using blood tests and radiographs of the chest	CatBoost,, SVM , LR	5148 patients were included in a retrospective cohort.	basic laboratory markers	89% to 96%
21	Patel, D., et al., [31]	COVID-19 disease severity predictors	Random Forest classifier	IRB	blood panel profile data	0.82
22	AlJame, M., et al., [32]	COVID-19 Diagnosis Using Routine	LightGBM, XGBoost, and extra trees	Kaggle	Biomarkers of the blood tests	99.5%

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 2267-2280
https://publishoa.com
ISSN: 1309-3452

		Blood Tests				
23	Kukar, M., et al., [33]	COVID-19 Diagnosis Using Routine Blood Tests	XGBoost	several high-dimensional datasets	Biomarkers of the blood tests	98%,
24	Qu, J., et al., [34]	The factors that influence the severity of COVID-19	logistic regression model	COVID-19 patients from seven Tokyo medical facilities	blood test results	80%
25	Zhang, J., et al., [35]	Individual COVID-19 diagnostic prediction	XGBoost, RF, LR	MSDW	Demographics , comorbidities, and common lab values.	79%
26	Gangloff, C., et al., [36]	machine learning based covid-19 diagnosis	LR, ANN, and RF	emergency department at Rennes Academic Hospital in France	Demographic and clinical data	92.4%
27	Hussain, L., et al., [37]	classifying COVID-19 lung infection	XGB-L, GXB-tree, CART, KNN, NB	Kaggle	Morphological features.	79% to 87%
28	Schmidt, M., et al., [38]	Predicting the survival of COVID-19 patients after 90 days.	XGBoost	SOSIC-14 dataset	Clinical Features	95%
29	Jaiswal, A. K., et al., [39]	chest x-rays and ct scans for the identification of covid-19	CNN	chest radiographs and computed tomography scans	Radiomic features	96%
30	Jha, S. K., et al., [40]	efficient disease diagnosis	fuzzy-rough-k-nearest neighbor	Kappa coefficient	efficient feature generation, selection, and classification methods	92%
31	Gomathi, S., et al., [41]	Predicting Covid-19 Pandemic	DT, NB, KNN, LR	Kaggle	Feature customization	95%
32	He, F., et al., [42]	COVID-19: Multicenter, Retrospective Study	XGBoost	electronic health record (EHR)	clinical features	95%
33	Monaghan, C.	Hemodialysis	XGBoost	Python	global	100%

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 2267-2280
https://publishoa.com
ISSN: 1309-3452

	K., et al., [43]	Prediction of patients with a SARS-CoV-2 illness that has gone undetected			importance of feature	
34	Blair, P. W., et al., [44]	HemodialysPrediction	machine learning, Topological data analysis	Kruskal-Wallis	Image features	82%

3. Observations and Findings

Because of user-friendly frameworks that provide outstanding results, advances in computer science have transformed how artificial intelligence is applied in academia, making Machine Learning (ML) methodologies accessible to researchers from a wide range of subjects. True, several current tendencies in the machine learning community are focused on winning rather than learning. Poor method selection justification leads to a disregard for the technique's limits, jeopardizing the translation of solutions into real-world clinical settings.

Since its initial breakout, COVID-19 has sparked a global pandemic. It needs to be under control as soon as possible. The growing number of instances around the world emphasizes the importance of quick, scalable, and reliable testing. Due to their slow reaction times, limited availability, and occasional inaccurate results, current diagnostic tests are difficult to use. The reverse transcription-polymerase chain reaction (RT-PCR) is the standard method for confirming COVID-19. To correct RT-false PCR's negative results, clinical, biochemical, and imaging data must be examined, just like any other test. Combining RT-PCR with chest-CT for patients with suspected COVID-19 could improve diagnostic performance, but this would need a significant investment in resources. In this example, machine learning has been researched in depth.

Patients who are at high risk of developing severe COVID-19 should take proactive measures to enhance their prognosis and utilize available medical resources. It is our obligation to determine what causes COVID-19 death in order to develop effective containment techniques and safeguard individuals who are most vulnerable.

As a first-line triage procedure, chest X-rays (CXRs) can be performed to screen non-COVID-19 patients with pneumonia. COVID-19 and pneumonia caused by other infections seem very similar on CXR pictures, making it difficult for radiologists to distinguish between the two. More variables and criteria are needed to accurately estimate the probability of infection from chest radiographs (CXR), which have a low sensitivity to the virus's effects.

Chest X-rays (CXRs) have been an important diagnostic tool in the COVID-19 disease diagnosis on a distinct level. The analysis of these photos largely relied on DL approaches. A single-scale benchmarking feature cannot capture the precise semantic information of infected lung areas due to variances in image resolution.

Patients with Ulcerative colitis or crohn's disease experience long-term digestive tract inflammation. Immunosuppressive medication for IBD can raise the risk of infection and make patients' symptoms worse. The effect of clinical and demographic variables on COVID-19 prognosis in patients with inflammatory bowel disease (IBD) is still being investigated. Progress has been hampered by a lack of data on a large number of COVID-19-infected IBD patients. ARDS (Acute respiratory distress syndrome) is a frequent condition in COVID-19 patients, and it is a major contributor to the current global ventilator shortage. The future research shall focus on clinical characteristics of COVID-19 in ARDS patients to enable the machine learning techniques to predict the risk of ARDS in COVID-19 patients.

COVID-19 might have produced precise estimates of COVID-19's unfavourable effects, enabling for more effective resource allocation and the execution of focused preventive actions. Despite the fact that COVID-19's epidemiology and clinical characteristics have been published, the risk factors that lead to severe disease in patients are unknown. Clinical outcomes can be predicted using machine learning algorithms that are supervised by professionals in the field. The underlying data distributions that constitute episodes, as well as the relationship between episode features and associated clinical outcomes, might alter over time due to the dynamic nature of clinical environments (data drift).

COVID-19 is a complex epidemiological system with many variables that make it challenging to forecast. Mathematical epidemiology strategies are the methods, which have been extensively using in forecasting epidemiology sequences and their impacts. However, most of these models relies on assumptions and tailored to specific scenarios. It is challenging to model a quickly moving epidemiological process with regional heterogeneity using commonly utilized compartmental models. The paucity of training data at the start of a pandemic, on the other hand, limits machine learning algorithms.

COVID-19 pre-screening utilizing crowdsourced cough sounds is another interesting discovery of the review. COVID-19 can be correctly identified from sound datasets, which accounts for the comparison of coughs of limited incidence of COVID-positive patients with healthy individuals as well as the number of different coughs for each recording.

Acute kidney damage (AKI) is a typical COVID-19 adverse effect that is used to determine the severity of the condition. Another significant goal of future research is to predict in-hospital mortality of the patients with COVID-19 and AKI.

Based on our findings, we believe that using patient EHR records to train machine learning (ML) techniques will enable us to better forecast and manage the risks associated with COVID-19 patient outcomes. Thousands of lives could be saved if SARS-CoV-2 epitope-blocking peptides or antibody sequences could be discovered quickly. The creation of high-throughput models for the prediction of SARS-CoV-2 neutralizing antibodies is one of several significant targets of current research. To anticipate drug-virus connection for drug repositioning, computational models based on matrix completion techniques must be created.

4. Conclusion

We compiled the most recent COVID-19 literature that centric to machine learning methods to track, contain, and treat viral infection in this systematic review. Our research sheds light on the prospect of machine learning in Covid-19 disease diagnosis, severity estimation and novel feature selection. Using machine learning applications, our research sheds light on the aforementioned contexts of the covid-19. As a result of our research, we believe that ML-based applications are critical for dealing with the wide range of Covid-19 contexts, such as epidemiology, disease diagnosis, and estimation of disease severity, drug and treatment analysis, and the estimation of post-Covid consequences.

References

[1]. Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J., & Hsueh, P. R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. International journal of antimicrobial agents, 55(3), 105924.

[2]. Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., ...& Feng, Z. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. New England journal of medicine.

[3]. Sun, J., Chen, X., Zhang, Z., Lai, S., Zhao, B., Liu, H., ...& Zheng, Y. (2020). Forecasting the long-term trend of COVID-19 epidemic using a dynamic model. Scientific reports, 10(1), 1-10.

[4]. Eurosurveillance Editorial Team. (2020). Note from the editors: World Health Organization declares novel coronavirus (2019-nCoV) sixth public health emergency of international concern. Eurosurveillance, 25(5), 200131e.

[5]. Ahmad, A., Garhwal, S., Ray, S. K., Kumar, G., Malebary, S. J., &Barukab, O. M. (2021). The number of confirmed cases of covid-19 by using machine learning: Methods and challenges. Archives of Computational Methods in Engineering, 28(4), 2645-2653.
[6]. Liu, G., &Rusling, J. F. (2021). COVID-19 antibody tests and their limitations. ACS sensors, 6(3), 593-612.
[7]. Druss, B. G. (2020). Addressing the COVID-19 pandemic in populations with serious mental illness. JAMA psychiatry, 77(9), 891-892.
[8]. Gibbons, R. C., Magee, M., Goett, H., Murrett, J., Genninger, J., Mendez, K., ...& Costantino, T. G. (2021). Lung ultrasound vs. chest X-ray study for the radiographic diagnosis of COVID-19 pneumonia in a high-prevalence population. The Journal of emergency medicine, 60(5), 615-625.
[9]. ZargariKhuzani, A., Heidari, M., &Shariati, S. A. (2021). COVID-Classifier: An automated machine learning model to assist in the diagnosis of COVID-19 infection in chest x-ray images. Scientific Reports, 11(1), 1-6.
[10]. Sotoudeh, H., Tabatabaei, M., Tasorian, B., Tavakol, K., Sotoudeh, E., &Moini, A. L. (2020). Artificial intelligence empowers radiologists to differentiate pneumonia induced by COVID-19 versus influenza viruses. ActaInformaticaMedica, 28(3), 190.
[11]. Awal, M. A., Masud, M., Hossain, M. S., Bulbul, A. A. M., Mahmud, S. H., &Bairagi, A. K. (2021). A novel bayesian optimization-based machine learning framework for COVID-19 detection from inpatient facility data. IEEE Access, 9, 10263-10281.
[12]. deFátimaCobre, A., Stremel, D. P., Noleto, G. R., Fachi, M. M., Surek, M., Wiens, A., ... &Pontarolo, R. (2021). Diagnosis and prediction of COVID-19 severity: can biochemical tests and machine learning be used as prognostic indicators?. Computers in biology and medicine, 104531.
[13]. Lorenzen, S. S., Nielsen, M., Jimenez-Solem, E., Petersen, T. S., Perner, A., Thorsen-Meyer, H. C., ...&Sillesen, M. (2021). Using machine learning for predicting intensive care unit resource use during the COVID-19 pandemic in Denmark. Scientific reports, 11(1), 1-10.
[14]. Kar, S., Chawla, R., Haranath, S. P., Ramasubban, S., Ramakrishnan, N., Vaishya, R., ...& Reddy, S. (2021). Multivariable mortality risk prediction using machine learning for COVID-19 patients at admission (AICOVID). Scientific reports, 11(1), 1-11.
[15]. Dabbah, M. A., Reed, A. B., Booth, A. T., Yassaee, A., Despotovic, A., Klasmer, B., ...& Mohan, D. (2021). Machine learning approach to dynamic risk modeling of mortality in COVID-19: a UK Biobank study. arXiv preprint arXiv:2104.09226.
[16]. Nakamichi, K., Shen, J. Z., Lee, C. S., Lee, A., Roberts, E. A., Simonson, P. D., ... & Van Gelder, R. N. (2021). Hospitalization and mortality associated with SARS-CoV-2 viral clades in COVID-19. Scientific reports, 11(1), 1-11.
[17]. Burdick, H., Lam, C., Mataraso, S., Siefkas, A., Braden, G., Dellinger, R. P., ...& Das, R. (2020). Prediction of respiratory decompensation in Covid-19 patients using machine learning: The READY trial. Computers in biology and medicine, 124, 103949.
[18]. Nguyen, S., Chan, R., Cadena, J., Soper, B., Kiszka, P., Womack, L., ...& Ray, P. (2021). Budget constrained machine learning for early prediction of adverse outcomes for COVID-19 patients. Scientific Reports, 11(1), 1-14.
[19]. Casiraghi, E., Malchiodi, D., Trucco, G., Frasca, M., Cappelletti, L., Fontana, T., ...&Valentini, G. (2020). Explainable machine learning for early assessment of COVID-19 risk prediction in emergency departments. IEEE Access, 8, 196299-196325.
[20]. Boussen, S., Cordier, P. Y., Malet, A., Simeone, P., Cataldi, S., Vaisse, C., ...&Bruder, N. (2021). Triage and monitoring of COVID-19 patients in intensive care using unsupervised machine learning. Computers in biology and medicine, 105192.
[21]. Ponce, D., Andrade, L. G. M., Granado, R. C., Ferrero, A., & Lombardi, R. (2021). Development of a Prediction Score for In-Hospital Mortality in COVID-19 Patients with Acute Kidney Injury: A Machine Learning Approach. Latin American Investigators AKI COVID-19 Registry, Development of a Prediction Score for In-Hospital Mortality in COVID-19 Patients with Acute Kidney Injury: A Machine Learning Approach.

[22]. Rechtman, E., Curtin, P., Navarro, E., Nirenberg, S., & Horton, M. K. (2020). Vital signs assessed in initial clinical encounters predict COVID-19 mortality in an NYC hospital system. Scientific reports, 10(1), 1-6.
[23]. Xu, W., Sun, N. N., Gao, H. N., Chen, Z. Y., Yang, Y., Ju, B., & Tang, L. L. (2021). Risk factors analysis of COVID-19 patients with ARDS and prediction based on machine learning. Scientific reports, 11(1), 1-12.
[24]. Estiri, H., Strasser, Z. H., & Murphy, S. N. (2021). Individualized prediction of COVID-19 adverse outcomes with MLHO. Scientific reports, 11(1), 1-9.
[25]. Roy, S., Sheikh, S. Z., &Furey, T. S. (2021). A machine learning approach identifies 5-ASA and ulcerative colitis as being linked with higher COVID-19 mortality in patients with IBD. Scientific reports, 11(1), 1-13.
[26]. Sayed, S. A. F., Elkorany, A. M., & Mohammad, S. S. (2021). Applying Different Machine Learning Techniques for Prediction of COVID-19 Severity. Ieee Access, 9, 135697-135707.
[27]. Duckworth, C., Chmiel, F. P., Burns, D. K., Zlatev, Z. D., White, N. M., Daniels, T. W., ...& Boniface, M. J. (2021). Using explainable machine learning to characterise data drift and detect emergent health risks for emergency department admissions during COVID-19. Scientific reports, 11(1), 1-10.
[28]. Heldt, F. S., Vizcaychipi, M. P., Peacock, S., Cinelli, M., McLachlan, L., Andreotti, F., ...& Khan, R. T. (2021). Early risk assessment for COVID-19 patients from emergency department data using machine learning. Scientific reports, 11(1), 1-13.
[29]. Murri, R., Lenkowicz, J., Masciocchi, C., Iacomini, C., Fantoni, M., Damiani, A., ...&Valentini, V. (2021). A Machine-learning Parsimonious Multivariable Predictive Model of Mortality Risk in Patients With Covid-19.
[30]. Du, R., Tsougenis, E. D., Ho, J. W., Chan, J. K., Chiu, K. W., Fang, B. X., ... &Vardhanabhuti, V. (2021). Machine learning application for the prediction of SARS-CoV-2 infection using blood tests and chest radiograph. Scientific reports, 11(1), 1-13.
[31]. Patel, D., Kher, V., Desai, B., Lei, X., Cen, S., Nanda, N., ...&Oberai, A. A. (2021). Machine learning based predictors for COVID-19 disease severity. Scientific Reports, 11(1), 1-7.
[32]. AlJame, M., Imtiaz, A., Ahmad, I., & Mohammed, A. (2021). Deep Forest Model for Diagnosing COVID-19 From Routine Blood Tests.
[33]. Kukar, M., Gunčar, G., Vovko, T., Podnar, S., Černelč, P., Brvar, M., ...&Notar, M. (2021). COVID-19 diagnosis by routine blood tests using machine learning. Scientific reports, 11(1), 1-9.
[34]. Qu, J., Sumali, B., Lee, H., Terai, H., Ishii, M., Fukunaga, K., ...& Nishimura, T. (2021). Finding of the factors affecting the severity of COVID-19 based on mathematical models. Scientific reports, 11(1), 1-7.
[35]. Zhang, J., Jun, T., Frank, J., Nirenberg, S., Kovatch, P., & Huang, K. L. (2021). Prediction of individual COVID-19 diagnosis using baseline demographics and lab data. Scientific Reports, 11(1), 1-8.
[36]. Gangloff, C., Rafi, S., Bouzillé, G., Soulat, L., &Cuggia, M. (2021). Author Correction: Machine learning is the key to diagnose COVID-19: a proof-of-concept study. Scientific Reports, 11(1), 1-1.
[37]. Hussain, L., Nguyen, T., Li, H., Abbasi, A. A., Lone, K. J., Zhao, Z., ...& Duong, T. Q. (2020). Machine-learning classification of texture features of portable chest X-ray accurately classifies COVID-19 lung infection. BioMedical Engineering OnLine, 19(1), 1-18.
[38]. Schmidt, M., Guidet, B., Demoule, A., Ponnaiah, M., Fartoukh, M., Puybasset, L., ...&Hajage, D. (2021). Predicting 90-day survival of patients with COVID-19: Survival of Severely Ill COVID (SOSIC) scores. Annals of intensive care, 11(1), 1-15.
[39]. Jaiswal, A. K., Tiwari, P., Rathi, V. K., Qian, J., Pandey, H. M., & Albuquerque, V. H. C. (2020). Covidpen: A novel covid-19 detection model using chest x-rays and ct scans. Medrxiv.
[40]. Jha, S. K., Marina, N., Wang, J., & Ahmad, Z. A hybrid machine learning approach of fuzzy-rough-k-nearest neighbor, latent semantic analysis, and ranker search for efficient disease diagnosis. Journal of Intelligent & Fuzzy Systems, (Preprint), 1-16.
[41]. Gomathi, S., Kohli, R., Soni, M., Dhiman, G., & Nair, R. (2020). Pattern analysis: predicting COVID-19 pandemic in India using AutoML. World Journal of Engineering.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 2267-2280
https://publishoa.com
ISSN: 1309-3452

[42]. He, F., Page, J. H., Weinberg, K. R., & Mishra, A. (2021). Development and validation of simplified machine learning algorithms to predict prognosis of hospitalized COVID-19 patients: a multi-centre, retrospective study. Journal of medical Internet research.

[43]. Monaghan, C. K., Larkin, J. W., Chaudhuri, S., Han, H., Jiao, Y., Bermudez, K. M., ...& Maddux, F. W. (2021). Machine Learning for Prediction of Hemodialysis Patients with an Undetected SARS-CoV-2 Infection. Kidney360.

[44]. Blair, P. W., Brandsma, J., Epsi, N. J., Richard, S. A., Striegel, D., Chenoweth, J., ... & Clark, D. (2021, November). 438. Phenotypic Differences Between Distinct Immune Biomarker Clusters During the 'Hyperinflammatory'Middle-Phase of COVID-19. In Open Forum Infectious Diseases (Vol. 8, No. Supplement_1, pp. S320-S321). US: Oxford University Press.