

A comprehensive review of early detection of COVID-19 based on machine learning and deep learning models

Ali J Askar Al-Khafaji, Nilam Nur Amir Sjarif

Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

Article Info

Article history:

Received Nov 14, 2023

Revised Mar 13, 2024

Accepted Mar 15, 2024

Keywords:

Artificial intelligence

COVID-19

Deep learning

Machine learning

Reverse transcription

Polymerase chain reaction

ABSTRACT

This paper reviews the use of machine learning (ML) and deep learning (DL) for early coronavirus disease (COVID-19) detection, highlighting their potential to overcome the limitations of traditional diagnostic methods such as long processing times and high costs. We analyze studies applying ML and DL to imaging, clinical, and genomic data, assessing their performance in terms of accuracy, sensitivity, specificity, and efficiency. The review discusses the advantages, limitations, and challenges of these models, including data quality, generalizability, and ethical considerations. It also suggests future research directions for improving model efficacy, such as integrating multi-modal data and developing more interpretable models. This concise review serves as a guide for researchers, healthcare practitioners, and policymakers on the advancements and prospects of ML and DL in early COVID-19 detection, promoting further innovation and collaboration in this vital public health domain.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Ali J Askar Al-Khafaji

Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia

54100 Kuala Lumpur, Malaysia

Email: jal-khafaji@graduate.utm.my, aliasker2@gmail.com

1. INTRODUCTION

The outbreak of the novel coronavirus disease (COVID-19) in late 2019 marked the onset of a global pandemic that has since exerted unprecedented strain on healthcare systems worldwide and disrupted the very fabric of society. As the virus continued to spread across continents, the imperative for rapid, accurate, and early detection of infections became critical to controlling the pandemic. Early detection not only facilitates timely treatment for individuals but also plays a pivotal role in implementing quarantine measures and managing public health resources effectively. Traditional diagnostic methods, such as reverse transcription-polymerase chain reaction (RT-PCR), while accurate, face challenges in terms of scalability, speed, and resource dependency. These constraints underscore the necessity for innovative approaches that can complement existing diagnostic protocols and enhance early detection capabilities.

In this context, machine learning (ML) and deep learning (DL) models have emerged as powerful tools for augmenting the early detection of COVID-19. Leveraging vast datasets, these models are capable of uncovering patterns and insights that may elude conventional analytical methods. From analyzing radiological images to interpreting complex clinical data, ML and DL have the potential to revolutionize the early detection process, offering faster and potentially more accurate diagnostics. This paper provides a comprehensive review of the current landscape of ML and DL applications in the early detection of COVID-19. It evaluates various models, their methodologies, data sources, performance metrics, and the challenges they face. Through this review, we aim to synthesize the existing knowledge, identify gaps in the current research, and suggest directions for future work to optimize the use of ML and DL in combating the

ongoing pandemic. This contribution is intended for researchers, healthcare professionals, and policymakers seeking to understand the potential of artificial intelligence in enhancing the early detection of COVID-19 and preparing for future public health challenges.

2. METHOD

In recent years, machine learning (ML) has played a pivotal role in COVID-19 research. ML, a subset of artificial intelligence (AI), focuses on crafting systems that evolve through examples without explicit instructions [1]. Its popularity stems from its prowess in addressing a myriad of real-world challenges. Core ML methodologies include reinforcement learning, unsupervised, and supervised. Supervised learning thrives on pre-labeled data sets, commonly through classification or regression. Conversely, unsupervised methods aim to derive insights from unclassified data, extracting features and patterns. Techniques like clustering fall under this umbrella. Reinforcement learning hinges on trial and error, leveraging a reward-penalty mechanism during training.

A comprehensive review by Albahri *et al.* [2] delved into the methodologies for collecting SARS-CoV-2 data and ML-based classification, as presented in computational biology. Meanwhile, Swapnarekha *et al.* [3] conducted an exhaustive analysis of how ML, mathematics, and statistical methods are utilized for COVID-19 projections and health assessments. A recent review by Tayarani [4] showcased the diversity of artificial intelligence tools employed against the pandemic, spanning areas from medical imaging to drug and epidemiological research. This literature encompasses AI subsets like deep learning, neural networks, and evolutionary algorithms. The same study also highlighted available public COVID-19 datasets. Wu *et al.* [5] explored the use of big data in identifying and managing COVID-19 cases in China.

Electronic health records (EHRs) house a plethora of data types, from complex to nonlinear. ML offers tools to simplify data interpretation and decision-making processes when applied to these records [6]. When trained on extensive EHRs, machine learning systems can discern intricate data interrelationships, often outperforming human capacities in tasks like image analysis or trend spotting [7]. Integrating ML algorithms with human clinical expertise can potentiate outcomes beyond individual capacities, as emphasized by Chen and Asch [8]. Several key algorithms, such as support vector machine (SVM) [9], naive bayes, and random forest [10], among others, have been highlighted in COVID-19 prediction research. In Canada, an ML-driven project was deployed to predict potential COVID-19 cases, utilizing stochastic fractal search in conjunction with numerical methods for predicting severe outcomes [11]. Zoabi *et al.* [12] utilized a gradient-boosting machine, widely regarded as a benchmark in tabular data prediction, to develop accurate projections. They employed the XGBoost model, designed to handle missing data proactively, using the AUC metric and early stopping based on validation sets.

2.1. Studies on COVID-19 diagnoses using imaging

Recent studies have shown that CT scans, originally intended for non-COVID-19 reasons like pre-elective procedure examinations and nervous system screenings, can be instrumental in identifying COVID-19 infections [13]. Clinical data may display features that resemble positive COVID-19 cases in contexts where CT scans are utilized, especially in individuals with respiratory problems but who tested negative via RT-PCR. As previously mentioned, deep learning plays a crucial role in ongoing COVID-19 research, spanning various sectors including epidemiology. However, issues like overfitting and hyperparameter adjustments are frequent in deep learning models. To address these challenges, this study recommends a metaheuristic-driven deep COVID-19 detection for X-ray images using a modified AlexNet structure. Parameter optimization for this enhanced AlexNet is achieved through the strength pareto evolutionary algorithm-II (SPEA-II), and the method is tested on a dataset comprising COVID-19, pneumonia, and healthy subjects [14].

Das *et al.* [15] proposed a model for detecting COVID-19 through X-ray imaging, employing deep learning and attaining a classification accuracy of 97.40%. Hemdan *et al.* [16] introduced COVIDX-Net, an automated system for detecting COVID-19 in chest X-rays, using seven deep learning models. Two of these, DenseNet201 and VGG19, both achieved 90% accuracy. It is worth noting that only 50 chest X-rays were used for validation. Luz *et al.* [17] explored deep learning models using the COVIDx database, demonstrating the efficacy of their proposed model. Their Flat EfficientNet model achieved an impressive 93.9% accuracy and 96.8% sensitivity. As a critical task in diagnostic systems, segmentation plays an essential role in accurately gauging COVID-19 infections in CT scans. A novel DL technique for the segmentation of various COVID-19 infection regions was introduced by Chen *et al.* [18], employing aggregated residual transformations alongside a soft learning algorithm to optimize model efficiency. Their research underscores the potential of DL-based segmentation in CT scan-based diagnosis of COVID-19. Jin *et al.* [19] proposed an AI-driven technique for rapid COVID-19 diagnosis and performed extensive CT

evaluations based on AI. Their CNN model achieved remarkable AUC scores when tested on large and diverse datasets. Civit-Masot *et al.* [20] explored a VGG16-based DL model for diagnosing pneumonia and COVID-19 using thoracic imaging, achieving 100% sensitivity for COVID-19 detection.

Various supervised machine learning methods have been employed to develop disease prediction models, with model performance sometimes varying based on the dataset. Several research initiatives aim to predict the early onset of this disease. Rangarajan and Ramachandran [21] used publicly available X-ray images and blood sample results to expedite COVID-19 prediction, testing their top-performing model on mobile devices. Zhou *et al.* [22] designed a predictive model for assessing COVID-19 severity using data from 377 patients, achieving commendable accuracy metrics. Their findings highlight three vital factors associated with COVID-19 severity. In a separate study by Luo *et al.* [23], a model was developed for early detection of mortality risk in COVID-19 cases, showcasing impressive results when certain biomarkers were combined. El Shenbary *et al.* [24] introduces a hybrid approach for feature extraction in COVID-19 image classification, combining traditional methods with deep learning (DL) neural networks. The method is evaluated on two distinct COVID-19 image datasets, demonstrating strong performance.

2.2. Studies on diagnosing COVID-19 based on blood and/or laboratory tests

Machine learning has been recently employed to forecast the mortality of COVID-19 patients, leading to the development of an XGBoost model. This model successfully predicted patient deaths approximately 10 days in advance with an accuracy exceeding 90% [25]. Using a convolutional neural network (CNN) and data from a Brazilian hospital, a model was developed to predict COVID-19, achieving an accuracy of 76% [26]. Another study aimed to refine COVID-19 severity predictions by examining blood test results. It revealed that the most influential factors determining severity were age, white blood cells, lymphocytes, and neutrophils [27]. In a study by Bayat *et al.* [28], a model was formulated to predict COVID-19 presence based on standard laboratory tests. Out of 75,991 participants provided by the US Department of Veterans Affairs, 7,335 tested positive. Using an XGBoost-based model, the study achieved 86.4% specificity, accuracy, and 82.4% sensitivity. The most significant parameters, in order of importance, were identified as temperature, white blood cell count, C-reactive protein, eosinophil count, serum ferritin, monocyte percentage, basophil count, dimer, and serum aspartate aminotransferase. Wu *et al.* [29] designed a classification tool using the random forest (RF) algorithm to promptly diagnose COVID-19 patients based on vital blood parameters. Using data from 169 potential COVID-19 cases from China, they built a model with 49 features. Eventually, 11 key input parameters were selected, with total bilirubin emerging as the most significant. Some less explored indicators, like total protein content, platelet distribution width, and basophils, showed potential as future diagnostic indicators. The model exhibited an AUC of nearly 99%, a sensitivity of 100%, and a specificity exceeding 94% on a separate test set.

In research by Tschoellitsch *et al.* [30], an RF-based ML model was developed using data from 1,528 patients to predict COVID-19 diagnosis. The model achieved 81% accuracy, 74% AUC, 60% sensitivity, and about 82% specificity. It identified leukocyte count, hemoglobin, red blood cell distribution width, and serum calcium as vital diagnostic parameters. Another study by Brinati *et al.* [31] employed various ML methods to predict COVID-19 diagnosis using data from 279 patients in Italy. The RF predictor was found to be the most effective, with an AUC and accuracy above 80%, a sensitivity close to 90%, and a specificity of 65%. Key factors for prediction included age, lymphocytes, white blood count (WBC), lactate dehydrogenase (LDH), alanine transaminase (ALT), eosinophils, and aspartate aminotransferase (AST). In a study by Tordjman *et al.* [32], 605 patients from four French hospitals participated. The research employed a logistic regression (LR) approach with a binary outcome to craft a scoring model predicting COVID-19 positivity. This model displayed an AUC of roughly 89%, an 80% sensitivity, and a 92% accuracy rate. Key biological markers such as lymphocytes, eosinophils, basophils, and neutrophils were highlighted in relation to COVID-19 diagnosis. Alakus and Turkoglu [33] presented a COVID-19 detection model built on an 18-feature dataset from 600 cases at Hospital Israelita Albert Einstein, Brazil. This dataset showcased diagnostic test results like hematocrit, hemoglobin, and platelet count. The research evaluated six deep learning architectures, with long short term memory (LSTM) emerging as the top performer, boasting an 86.66% accuracy, 99.42% recall, and 62.50% AUC. In a study by Batista *et al.* [34], a model was developed using 235 COVID patient data, including 102 positive samples. Various techniques such as SVM, gradient boosting decision trees (GBDT), nearest neighbor algorithm (NN), random forest (RF), and LR were explored. The SVM emerged as the superior choice, delivering an AUC of 85%, sensitivity of 68%, and specificity of 85%. Lymphocyte, leukocyte, and eosinophil counts were identified as crucial diagnostic predictors.

2.3. Epidemiology-informed diagnoses of COVID-19 research

Liang *et al.* [35] designed a predictive tool to determine individuals at higher risk of developing severe health conditions. Their study included 2,300 participants from 575 medical institutions, where 1,590

participants were used to design the model, and 710 were chosen for its validation. Using least absolute shrinkage and selection operator (LASSO) regression, they sifted through 72 potential predictors and narrowed them down to 19 for linear regression (LR) risk prediction models. Ultimately, only 10 key criteria were chosen for the vulnerability index: Chest radiography (CXR) abnormalities, age, presence of hemoptysis, drowsiness, number of underlying conditions, history of LDH, neutrophil to lymphocyte ratio, cancer, and direct bilirubin levels. The risk evaluation was based on these factors, and the model's accuracy reached 0.88 for both the calibration and validation sets.

Li *et al.* [36] investigated the potential correlation between the mortality risk of COVID-19 patients and various factors, including underlying health conditions, age, and gender. Two datasets were used: the GitHub dataset with 28,958 cases (and 530 post-treatment fatalities) and the Wolfram dataset with 1,448 cases (and 123 deaths after data preprocessing). The analysis found that results from the Wolfram dataset were more insightful than the GitHub one, possibly due to the latter's lack of granular data and consistent consensus per case. Multiple machine learning and data analysis techniques were employed, such as RF, LR, SVM, isolation forest, one-class SVM, autoencoder, and local outlier factor. The autoencoder model delivered the most promising outcomes, with an AUC of about 73%, accuracy at 97.5%, specificity at 97.5%, and a sensitivity of 40.05%. The findings highlighted a significant association between mortality and chronic illnesses or symptoms related to digestive, kidney, heart, or respiratory ailments.

2.4. Studies focusing on COVID-19 diagnosis through symptoms

In 2021 research by Kukar *et al.* [37], deep neural networks (DNN) and XGBoost methodologies were employed. These algorithms, utilizing gender, blood test results, and age, aimed to predict future COVID-19 cases. Within the study's 5333 participants from the University Medical Center Ljubljana, 160 tested positive for COVID-19. XGBoost emerged as the most effective, boasting an 82% sensitivity, 97% AUC, and a specificity nearing 98%. Key indicators for COVID-19 identification in the model included albumin levels, prothrombin activity ratio, and mean corpuscular hemoglobin concentration (MCHC), among others. Importantly, distinctions between COVID-19 and bacterial infections were driven by metrics like leukocyte count and hemoglobin levels. Several factors, including an individual's prothrombin activity, played pivotal roles in distinguishing COVID-19 from other infections.

Shoer *et al.* [38] developed a predictive model rooted in nine basic survey questions. Their dataset, sourced from an expansive symptom survey taken by over two million Israelis, spotlighted 43,752 participants, 498 of whom tested positive for the virus. The survey delved into demographic data and self-reported symptoms. Utilizing logistic regression, the research achieved an AUC of 0.737. In the study by Izquierdo *et al.* [39], a strategy to predict intensive care unit (ICU) admissions was hatched using machine learning. The study captured data from 10,504 COVID-19 patients from Castilla La Mancha, Spain. Of these, 1,353 were hospitalized and 83 were admitted to the ICU. Decision trees (DT) were selected as the analytical approach, yielding precision, recall, and AUC values within the 0.68 to 0.76 range. Factors like age and fever frequency influenced ICU admissions.

Gao *et al.* [6] posited that patients' clinical data at the point of admission could predict mortality risks for up to 20 days. Their sample comprised 251 patients from Wuhan's Central Hospital, associated with Tongji Medical College. Multiple methodologies, including neural networks and logistic regression, were explored. The study highlighted several factors positively or negatively correlated with mortality. Testing across three cohorts yielded AUC values ranging from 92.46% to 97.60%. In the study by Suma *et al.* [40], machine learning algorithms were crafted from 65,000 clinical records encompassing 26 features. An optimized feature set was pinpointed using a refined ABC optimization technique. The SVM, once applied, proved most precise at 96% accuracy. The research also adeptly predicted the severity of the disease in COVID-positive patients. Deniz *et al.* [41] proposed a multi-threaded genetic feature selection algorithm integrated with extreme learning machines (MG-ELM) to predict COVID-19 severity. The dataset, which underwent refinement from 15 to 34 features, initially contained redundancies and missing values. The updated dataset emphasized 24 symptom features. The study's outcomes revealed that the MG-ELM outperformed other machine learning methodologies in accuracy.

On the other hand, the primary objectives of Solayman *et al.* [42] research was to forecast whether individuals are infected with the coronavirus and raise awareness about their COVID-19 situations, aiming to contribute to the prevention of future disease spread. The study leverages an open-source dataset containing information on over 2 million individuals, encompassing their symptoms and crucial details such as test dates, results, gender, and age. Various data preprocessing methods, including addressing null values, converting categorical features, and handling imbalanced datasets using the synthetic minority over-sampling technique (SMOTE) approach, are employed. Subsequently, multiple machine learning approaches, accompanied by hyperparameter tuning, are implemented. Among these, the CNN-LSTM model, combined with the SMOTE approach, achieves the most accurate predictions in terms of classification accuracy and F1

score. To interpret the prediction outcomes, an explainable AI technique utilizing the LIME framework is applied. Finally, the developed machine learning model is deployed on a website. One notable limitation of the study is the use of an open-source data set specific to a particular region's patient data. Future research endeavors could benefit from utilizing a private dataset with additional biomarkers, covering a broader range of features and regions. To enhance prediction accuracy, the incorporation of meta-learning techniques and a fusion of machine learning models with fuzzy logic frameworks are suggested. Additionally, feature selection using the wrapper technique could be implemented to boost the system's performance.

El Massari *et al.* [43] in their study introduced and evaluated seven machine learning (ML) algorithms alongside an ontology-based model, including a detailed performance comparison. It utilized two testing approaches: 10-fold cross-validation and percentage split, applying various metrics like accuracy, F-measure, precision, and recall to gauge the effectiveness of each method. The results highlighted the ontology model's superior accuracy, even without the implementation of feature selection. This discovery opens up a new avenue for research, encouraging scholars to delve into this area to uncover further insights and contribute additional findings. Such contributions could enhance forecasting, recommendations, or decision-making processes. Looking ahead, this paper aims to refine their comparative analysis by exploring innovative strategies for integrating ML rules with ontology-based methods and incorporating regression ML algorithms. Table 1 shows the comparison of certain studies about COVID-19 diagnosis through Symptoms, including the method these studies used, the quantity of variables, the number of cases included in the datasets, the evaluation of performance and the results obtained from using these techniques.

3. RESULTS AND DISCUSSION

The extensive review of current research on COVID-19 diagnosis methods reveals significant advancements in employing machine learning (ML), artificial intelligence (AI), and various data analysis techniques to combat the pandemic. Key findings from the literature indicate:

- Machine learning and AI efficacy: ML and AI have demonstrated substantial potential in diagnosing COVID-19, predicting disease severity, and forecasting spread patterns. Techniques such as supervised learning, unsupervised learning, deep learning, and reinforcement learning have been applied across different stages of COVID-19 management, from early detection using imaging and blood tests to predicting patient outcomes and mortality rates.
- Data sources and methodologies: The studies utilized diverse data sources, including electronic health records (EHRs), imaging data (CT scans and X-rays), blood tests, and demographic studies. The application of ML algorithms like SVM, naive Bayes, random forest, XGBoost, and deep learning models (e.g., CNNs) has been highlighted. These methods have been crucial in extracting meaningful insights from complex and high-dimensional data.
- Predictive accuracy and model performance: The predictive models have achieved notable accuracy in diagnosing COVID-19, assessing disease severity, and projecting future cases. Models based on imaging data, particularly CT scans and X-rays, have shown high efficacy in detecting COVID-19 infections. Similarly, models analyzing blood tests and demographic data have been effective in predicting the disease's presence and outcomes.
- Innovations and challenges: Research has introduced innovative approaches, such as the use of blockchain for data sharing, incentive-based methods for disease containment, and hybrid models combining traditional algorithms with deep learning. However, challenges remain, such as data quality, model overfitting, the need for large and diverse datasets, and the integration of ML models with clinical decision-making processes.

4. CONCLUSION

Throughout the current pandemic, there's been an influx of research dedicated to halting its progression. These initiatives delve into varied fields, from medical imaging and blood test outcomes to the epidemiological backgrounds of patients. A convergence of these techniques is also noticeable, encompassing demographic data for a comprehensive view. Even though RT-PCR stands as a chief detection method, its potential error rate, reaching up to 33%, calls for supplementary diagnostic strategies.

A review of the literature highlights the variety of datasets used in COVID-19 detection. However, hurdles remain. Concerns about data confidentiality have limited the accessibility of essential datasets, adding complexities to studies in this area. In addition, while tools like X-Rays and CT-scans offer important data, their repeated usage raises concerns over health implications. Thus, radiological methods are not as preferred due to health concerns from repeated use. Similarly, blood tests face challenges; the limited availability of samples and the requirement for larger datasets for ML approaches make them a less practical

solution. The vast number of infected individuals and the trend of self-isolation outside formal health systems further challenge the thorough monitoring of cases. Here are the main characteristics of the work we've conducted:

- COVID-19 symptoms-based datasets and their derived features have not been thoroughly examined for detecting infected cases. For instance, the SIVEP-Gripe dataset remains under-researched in the context of symptom-based detection.
- A surge in false negative results can have significant repercussions, often stemming from the fact that many extracted features contain redundant, irrelevant, or noisy data.
- The exploration of evolutionary feature selection in COVID-19 symptom-based datasets remains limited.
- COVID-19 datasets sourced from various origins tend to be imbalanced, potentially impacting classification outcomes and causing skewed model evaluation results.

APPENDIX

Table 1. Studies focusing on COVID-19 diagnosis through symptoms (*continues*)

References	Method	Variables	Dataset	Evaluation performance	Results
Li <i>et al.</i> [44]	XGBoost	27	413	ROC curve, PR curve, and AUC for the two curves.	The XGBoost model yielded a sensitivity of 92.5 and a specificity of 97.9.
Shoer <i>et al.</i> [38]	Logistic Regression	9	43752	Four folds cross-validation	The model in this study yielded results of 0.737 for auROC and 0.144 for auPR.
Cabitzza <i>et al.</i> [45]	The recursive feature-elimination (RFE) algorithm, SVM, and KNN were employed alongside RF, NB, LR, and SVM.	72	1624	Five folds cross-validation and AUC	In this study, we developed five distinct machine learning models. The first model focuses on completing the OSR dataset; the second targets the AUC; the third spans algorithm ranges from 0.83 to 0.90; the fourth is tailored to the COVID-specific dataset with a range of 0.83 to 0.87; and the last is designed for the CBC dataset ranging from 0.74 to 0.86. Furthermore, the validation results yielded impressive scores, with AUC values between 0.75 and 0.78 and specificity values ranging from 0.92 to 0.96.
Gao <i>et al.</i> [6]	SVM, GBDT, LASSO, LR and NN	34 to 8	251	ROC curve, (Kaplan–Meier) curve, (calibration) curve, and (evaluation metrics) including AUC). For validation this paper included ((Accuracy), (sensitivity), (specificity), (PPV), (NPV), (F1 score), (Cohen’s Kappa coefficient), and (Brier score)).	Of the 34 traits examined, 8 displayed a positive correlation with mortality, whereas 6 showed a negative link. The AUCs for the three testing cohorts were 96.21% for the internal validation and 97.60% and 92.46% for the two external validations, respectively.
Aljameel <i>et al.</i> [46]	XGB, RF, and LR.	20	287	SMOTE and (0-k cross-validation) sensitivity, AUC	RF was as high as (0.95) while AUC was (0.99).
Mei <i>et al.</i> [47]	MLP, random forest and SVM	8	COVID19-CT		AUC was (0.92).
Wagner <i>et al.</i> [48]	BERT-based neural networks	26	77167	(Precision, Accuracy and Recall)	Results showed that "Anosmia or dysgeusia" had a 27.1-fold increase, "Fever or chills" had a 2.6-fold increase, "Respiratory difficulties" and "Cough" each had a 2.2-fold increase, "Myalgia or arthralgia" recorded a 2.0-fold increase, and "Diarrhea" showed a 1.4-fold increase in amplifying COVID-19 positive cases compared to negative ones. When combined, cough and fever showed as high as a 4.2-fold increase. auROC was (0.90) and auPRC was (0.66).
Zoabi <i>et al.</i> [12]	a gradient-boosting model with DT	8	51831	False-positive rate, false-negative rate, false discovery rate, specificity, PPV and negative predictive value, auROC, sensitivity and accuracy	

Table 1. Studies focusing on COVID-19 diagnosis through symptoms (*continues*)

References	Method	Variables	Dataset	Evaluation performance	Results
Deniz <i>et al.</i> [41]	Multi-threaded (GA) with (MG-ELM)	34 to 10	1085	Accuracy	MG-ELM outperformed other ML methods, achieving an accuracy rate of 96.22%.




REFERENCES

- [1] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, "A survey of deep Learning and its applications: A new paradigm to machine learning," *Archives of Computational Methods in Engineering*, vol. 27, no. 4, pp. 1071–1092, Jun. 2020, doi: 10.1007/s11831-019-09344-w.
- [2] A. S. Albahri *et al.*, "Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (COVID-19): A systematic review," *Journal of Medical Systems*, vol. 44, no. 7, May 2020, doi: 10.1007/s10916-020-01582-x.
- [3] H. Swapnarekha, H. S. Behera, J. Nayak, and B. Naik, "Role of intelligent computing in COVID-19 prognosis: A state-of-the-art review," *Chaos, Solitons and Fractals*, vol. 138, p. 109947, Sep. 2020, doi: 10.1016/j.chaos.2020.109947.
- [4] M. H. Tayarani N., "Applications of artificial intelligence in battling against covid-19: A literature review," *Chaos, Solitons and Fractals*, vol. 142, p. 110338, Jan. 2021, doi: 10.1016/j.chaos.2020.110338.
- [5] J. Wu, J. Wang, S. Nicholas, E. Maitland, and Q. Fan, "Application of big data technology for COVID-19 prevention and control in China: Lessons and recommendations," *Journal of Medical Internet Research*, vol. 22, no. 10, p. e21980, Oct. 2020, doi: 10.2196/21980.
- [6] Y. Gao *et al.*, "Machine learning based early warning system enables accurate mortality risk prediction for COVID-19," *Nature Communications*, vol. 11, no. 1, p. 5033, Oct. 2020, doi: 10.1038/s41467-020-18684-2.
- [7] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, Apr. 2019, doi: 10.1056/nejmra1814259.
- [8] J. H. Chen and S. M. Asch, "Machine learning and prediction in medicine — beyond the peak of inflated expectations," *New England Journal of Medicine*, vol. 376, no. 26, pp. 2507–2509, Jun. 2017, doi: 10.1056/nejmp1702071.
- [9] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/bf00994018.
- [10] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998, doi: 10.1109/34.709601.
- [11] S. Khalilpourazari and H. Hashemi Doulabi, "Robust modelling and prediction of the COVID-19 pandemic in Canada," *International Journal of Production Research*, vol. 61, no. 24, pp. 8367–8383, Jun. 2023, doi: 10.1080/00207543.2021.1936261.
- [12] Y. Zoabi, S. Deri-Rozov, and N. Shomron, "Machine learning-based prediction of COVID-19 diagnosis based on symptoms," *npj Digital Medicine*, vol. 4, no. 1, Jan. 2021, doi: 10.1038/s41746-020-00372-6.
- [13] W. Zhao, W. Jiang, and X. Qiu, "Deep learning for COVID-19 detection based on CT images," *Scientific Reports*, vol. 11, no. 1, Jul. 2021, doi: 10.1038/s41598-021-93832-2.
- [14] M. Kaur, V. Kumar, V. Yadav, D. Singh, N. Kumar, and N. N. Das, "Metaheuristic-based deep COVID-19 screening model from chest X-Ray images," *Journal of Healthcare Engineering*, vol. 2021, pp. 1–9, Mar. 2021, doi: 10.1155/2021/8829829.
- [15] N. Narayan Das, N. Kumar, M. Kaur, V. Kumar, and D. Singh, "Automated deep transfer learning-based approach for detection of COVID-19 infection in chest X-rays," *Irbm*, vol. 43, no. 2, pp. 114–119, Apr. 2022, doi: 10.1016/j.irbm.2020.07.001.
- [16] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "COVIDX-Net: A framework of deep learning classifiers to diagnose COVID-19 in X-Ray images," *arxiv.org/abs/2003.11055*, Mar. 2020.
- [17] E. Luz *et al.*, "Towards an effective and efficient deep learning model for COVID-19 patterns detection in X-ray images," *Research on Biomedical Engineering*, vol. 38, no. 1, pp. 149–162, Apr. 2022, doi: 10.1007/s42600-021-00151-6.
- [18] X. Chen, L. Yao, and Y. Zhang, "Residual attention U-Net for automated multi-class segmentation of COVID-19 chest CT images," *arxiv.org/abs/2004.05645*, Apr. 2020.
- [19] C. Jin *et al.*, "Development and evaluation of an artificial intelligence system for COVID-19 diagnosis," *Nature Communications*, vol. 11, no. 1, Oct. 2020, doi: 10.1038/s41467-020-18685-1.
- [20] J. Civit-Masot, F. Luna-Perejón, M. D. Morales, and A. Civit, "Deep learning system for COVID-19 diagnosis aid using X-ray pulmonary images," *Applied Sciences (Switzerland)*, vol. 10, no. 13, p. 4640, Jul. 2020, doi: 10.3390/app10134640.
- [21] A. K. Rangarajan and H. K. Ramachandran, "A preliminary analysis of AI based smartphone application for diagnosis of COVID-19 using chest X-ray images," *Expert Systems with Applications*, vol. 183, Nov. 2021, doi: 10.1016/j.eswa.2021.115401.
- [22] Y. Zhou *et al.*, "A new predictor of disease severity in patients with COVID-19 in Wuhan, China," *medRxiv*, p. 2020.03.24.20042119, Mar. 2020, doi: 10.1101/2020.03.24.20042119.
- [23] M. Luo, J. Liu, W. Jiang, S. Yue, H. Liu, and S. Wei, "IL-6 and CD8+ T cell counts combined are an early predictor of in-hospital mortality of patients with COVID-19," *JCI Insight*, vol. 5, no. 13, Jul. 2020, doi: 10.1172/jci.insight.139024.
- [24] H. A. El Shenbary, E. A. Ebeid, and D. Baleanu, "COVID-19 classification using hybrid deep learning and standard feature extraction techniques," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 29, no. 3, pp. 1780–1791, Mar. 2023, doi: 10.11591/ijeecs.v29.i3.pp1780-1791.
- [25] L. Yan *et al.*, "Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan," *medRxiv*, Mar. 2020, doi: 10.1101/2020.02.27.20028027.
- [26] H. Turabieh and W. B. A. Karaa, "Predicting the existence of COVID-19 using machine learning based on laboratory findings," Mar. 2021, doi: 10.1109/WIDSTAI52235.2021.9430233.
- [27] J. Luo, L. Zhou, Y. Feng, B. Li, and S. Guo, "The selection of indicators from initial blood routine test results to improve the accuracy of early prediction of COVID-19 severity," *PLoS ONE*, vol. 16, no. 6 June, p. e0253329, Jun. 2021, doi: 10.1371/journal.pone.0253329.
- [28] V. Bayat *et al.*, "A severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) prediction model from standard laboratory tests," *Clinical Infectious Diseases*, vol. 73, no. 9, pp. E2901–E2907, Aug. 2021, doi: 10.1093/cid/ciaa1175.
- [29] J. Wu *et al.*, "Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results," *medRxiv*, p. 2020.04.02.20051136, Apr. 2020, doi: 10.1101/2020.04.02.20051136.
- [30] T. Tschoellitsch, M. Dünsen, C. Böck, K. Schwarzbauer, and J. Meier, "Machine learning prediction of SARS-CoV-2 polymerase chain reaction results with routine blood tests," *Lab Medicine*, vol. 52, no. 2, pp. 146–149, Dec. 2021, doi: 10.1093/labmed/lmaa111.




- [31] D. Brinati, A. Campagner, D. Ferrari, M. Locatelli, G. Banfi, and F. Cabitza, "Detection of COVID-19 infection from routine blood exams with machine learning: A feasibility study," *Journal of Medical Systems*, vol. 44, no. 8, Jul. 2020, doi: 10.1007/s10916-020-01597-4.
- [32] M. Tordjman *et al.*, "Pre-test probability for SARS-Cov-2-related infection score: The PARIS score," *PLoS ONE*, vol. 15, no. 12, December, p. e0243342, Dec. 2020, doi: 10.1371/journal.pone.0243342.
- [33] T. B. Alakus and I. Turkoglu, "Comparison of deep learning approaches to predict COVID-19 infection," *Chaos, Solitons and Fractals*, vol. 140, p. 110120, Nov. 2020, doi: 10.1016/j.chaos.2020.110120.
- [34] A. Batista *et al.*, "COVID-19 diagnosis prediction in emergency care patients: a machine learning approach," *medRxiv*, p. 2020.04.04.20052092, Apr. 2020, doi: 10.1101/2020.04.04.20052092.
- [35] W. Liang *et al.*, "Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19," *JAMA Internal Medicine*, vol. 180, no. 8, pp. 1081–1089, Aug. 2020, doi: 10.1001/jamainternmed.2020.2033.
- [36] Y. Li *et al.*, "Individual-level fatality prediction of COVID-19 patients using AI methods," *Frontiers in Public Health*, vol. 8, Sep. 2020, doi: 10.3389/fpubh.2020.587937.
- [37] M. Kukar *et al.*, "COVID-19 diagnosis by routine blood tests using machine learning," *Scientific Reports*, vol. 11, no. 1, May 2021, doi: 10.1038/s41598-021-90265-9.
- [38] S. Shoer *et al.*, "A prediction model to prioritize individuals for a SARS-CoV-2 test built from national symptom surveys," *Med*, vol. 2, no. 2, pp. 196–208.e4, Feb. 2021, doi: 10.1016/j.medj.2020.10.002.
- [39] J. L. Izquierdo, J. Ancochea, and J. B. Soriano, "Clinical characteristics and prognostic factors for intensive care unit admission of patients with COVID-19: Retrospective study using machine learning and natural language processing," *Journal of Medical Internet Research*, vol. 22, no. 10, p. e21801, Oct. 2020, doi: 10.2196/21801.
- [40] L. S. Suma, H. S. Anand, and S. S. Vinod chandra, "Nature inspired optimization model for classification and severity prediction in COVID-19 clinical dataset," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 3, pp. 1699–1711, Jul. 2023, doi: 10.1007/s12652-021-03389-1.
- [41] A. Deniz, H. E. Kiziloz, E. Sevinc, and T. Dokeroglu, "Predicting the severity of COVID-19 patients using a multi-threaded evolutionary feature selection algorithm," *Expert Systems*, vol. 39, no. 5, Feb. 2022, doi: 10.1111/exsy.12949.
- [42] S. Solayman, S. A. Aumi, C. S. Mery, M. Mubassir, and R. Khan, "Automatic COVID-19 prediction using explainable machine learning techniques," *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 36–46, Jun. 2023, doi: 10.1016/j.ijcce.2023.01.003.
- [43] H. El Massari, N. Gherabi, S. Mhammedi, H. Ghandi, F. Qanouni, and M. Bahaj, "Integration of ontology with machine learning to predict the presence of covid-19 based on symptoms," *arxiv.org/abs/1205.5923*, May 2012.
- [44] W. T. Li *et al.*, "Using machine learning of clinical data to diagnose COVID-19: A systematic review and meta-analysis," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, Sep. 2020, doi: 10.1186/s12911-020-01266-z.
- [45] F. Cabitza *et al.*, "Development, evaluation, and validation of machine learning models for COVID-19 detection based on routine blood tests," *Clinical Chemistry and Laboratory Medicine*, vol. 59, no. 2, pp. 421–431, Oct. 2021, doi: 10.1515/cclm-2020-1294.
- [46] S. S. Aljameel, I. U. Khan, N. Aslam, M. Aljabri, and E. S. Alsulmi, "Machine learning-based model to predict the disease severity and outcome in COVID-19 patients," *Scientific Programming*, vol. 2021, pp. 1–10, Apr. 2021, doi: 10.1155/2021/5587188.
- [47] X. Mei *et al.*, "Artificial intelligence-enabled rapid diagnosis of patients with COVID-19," *Nature Medicine*, vol. 26, no. 8, pp. 1224–1228, May 2020, doi: 10.1038/s41591-020-0931-3.
- [48] T. Wagner *et al.*, "Decision letter: Augmented curation of clinical notes from a massive EHR system reveals symptoms of impending COVID-19 diagnosis," *eLife Sciences Publications, Ltd*, Jun. 2020. doi: 10.7554/elife.58227.sa1.

BIOGRAPHIES OF AUTHORS



Ali J Askar Al-Khafaji    is currently doing PhD degree at Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia since 2020. He is currently holding bachelor's degree in software engineering in addition to master's degree from Pune university, India. His research areas are data mining, software engineering, machine learning and text mining. He can be contacted at emails: aliasker2@gmail.com, jal-khafaji@graduate.utm.my.



Nilam Nur Amir Sjarif    is a senior lecturer at Razak Faculty of Technology and Informatics (RFTI), Universiti Teknologi Malaysia. Currently, she is the quality coordinator at RFTI, UTM. She received her PhD from the Universiti Teknologi Malaysia, Skudai in 2015, in the area of human action recognition behavior for video surveillance. She is active in the research field of machine learning, deep learning, natural language processing, pattern recognition, image processing and big data analytics. Other than working in the academic world, she is also involved in several research projects with industries, such as JAKIM and Cybersecurity Malaysia and Maritime Institute of Malaysia (MIMA). She can be contacted at email: nilamnur@utm.my.