IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 28 April 2024, accepted 21 May 2024, date of publication 24 May 2024, date of current version 3 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3405071

RESEARCH ARTICLE

An Explainable Decision Support Framework for Differential Diagnosis Between Mild COVID-19 and Other Similar Influenzas

KRISHNARAJ CHADAGA[1], SRIKANTH PRABHU[1], (Senior Member, IEEE), NIRANJANA SAMPATHILA[2], (Senior Member, IEEE), RAJAGOPALA CHADAGA[3], AND SHASHIKIRAN UMAKANTH[4]

[1]Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India
[2]Department of Biomedical Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India
[3]Department of Mechanical and Industrial Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India
[4]Department of Medicine, Dr. TMA Pai Hospital, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

Corresponding authors: Srikanth Prabhu (srikanth.prabhu@manipal.edu) and Niranjana Sampathila (niranjana.s@manipal.edu)

ABSTRACT It is tough to clinically differentiate between mild COVID-19 and other similar influenzas due to their comparable transmission traits and symptoms. The Real-time reverse transcriptase-polymerase chain reaction (RT-PCR) test is utilized regularly to diagnose severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) despite being prone to false-negative results. In recent years, intelligent support systems have been developed for patient triage and disease diagnosis. Thus, this research utilizes machine learning to diagnose COVID-19 from routine biomarkers. Twelve feature selection techniques, which include nature-inspired techniques, have been compared to extract the essential features. Multiple classifiers, including stacking, voting and deep learning, are trained to predict the patient diagnosis. The maximum accuracy obtained by the classifiers was 95% in this retrospective study. The diagnostic predictions were further interpreted using five explainable artificial intelligence methods. Biomarkers such as albumin, protein, eosinophil and total white blood cells were crucial. Thus, automated diagnostic systems can be supportive in the accurate and timely detection of COVID-19 and similar influenza infections.

INDEX TERMS Biomarkers, explainable artificial intelligence, Mild COVID-19, nature-inspired algorithms, non-COVID-19 influenza.

I. INTRODUCTION

Mild COVID-19 symptoms include fever, myalgia, cough, sore throat, and other flu-like manifestations [1]. However, the above symptoms are also observed in other non-COVID-19 influenza [2]. Early diagnosis is crucial for preventing infections and lowering the fatality rate. Therefore, it is essential to establish effective diagnostic techniques to detect COVID-19 quickly. Due to several similarities between non-COVID-19 influenza and mild COVID-19, detection has become more complex. Because both infections exist alongside one another, it is essential to distinguish between them to treat individuals at a high infection risk, some of whom may have symptoms specific to the virus [3]. Some of the mild symptoms observed in patients are depicted in Fig. 1.

Two diagnostic methodologies, Real-time reverse transcriptase polymerase chain reaction (RT-PCR) and antibody testing, have been used for most patients [4]. However, both the above methods have their limitations. RT-PCR tests have been known to generate false-negative results [5]. Handling specimens and high costs have also been an issue [5]. Antibody testing showed an extremely low sensitivity, too [6]. Hence, other diagnostic techniques, such as imaging, voice-based detection and clinical markers, were also tested [7].

The healthcare sector has already made tremendous use of artificial intelligence (AI) and machine learning (ML)

The associate editor coordinating the review of this manuscript and approving it for publication was Ghulam Muhammad.

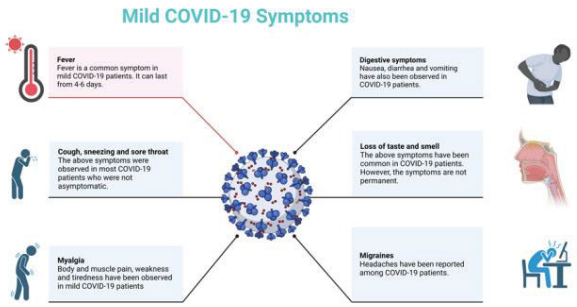

FIGURE 1. Mild symptoms observed in COVID-19 patients.

techniques. In the modern world, disease detection and prognosis are accurately predicted using soft computing techniques [8]. The doctor's decision-making ability can be improved using multiple cognitive and visualization techniques. The foregoing developments are primarily attributable to the rise of pertinent clinical datasets and technologies [9]. Before using automated AI systems in real-life situations, assessing their transparency and interpretability is vital. The reasoning behind a classifier's prediction must be clear to the medical personnel [10]. Models which are hard to understand cannot be used for real-time diagnosis. A new sub-branch of AI called explainable AI (XAI) makes the predictions transparent and trustworthy [11]. The algorithms' visualisation techniques make it easier for doctors to understand the diagnosis.

Blood and routine laboratory tests diagnose various diseases, including COVID-19. However, no single blood marker can detect this contagious virus. However, combining markers such as albumin, neutrophil to lymphocyte (NLR) ratio, eosinophils, neutrophils, lymphocytes, and other clinical markers have been helpful [12], [13]. This offers an unparalleled chance for investigators to examine characteristics affecting patient diagnosis while creating novel testing approaches for coronavirus detection. Some similar research conducted by other investigators are described below.

In a research by Altantawy and Kishk [14], a diagnostic classifier based on equilibrium methodology was utilized for detecting this virus. A 1D-CNN model was used on the San Raphael Hospital dataset. A testing accuracy of 98.50% was obtained in this study. Self-organizing maps and cluster analysis were utilized for COVID-19 detection in another study [15]. The best classifier obtained an accuracy of 83%. A technique called ''Information needs'' was used to identify essential markers. According to it, the crucial blood variables were Leukocytes and Basophils. Mohammadi et al. [16] utilized artificial neural networks to classify COVID-19 from routine laboratory markers. According to the research, the critical blood parameters were CRP, magnesium, vitamin D and lymphocyte count. The multilayer perceptron network obtained an accuracy of 99.16%. Another research used explainable learning for automatic COVID-19 detection [17]. The data was from Israel, and the model reached an accuracy of 96.34%. The research also utilized the ''Local Interpretable Model-agnostic Explanations (LIME)'' framework for demystifying the algorithms. Roland et al. [18] used a domain shifts-based method to diagnose the virus. The data was obtained from Johannes Kepler University, Austria. The best performing machine learning algorithm obtained an AUC of 0.88. A machine learning methodology was used for differential diagnosis between COVID-19 and other similar diseases in children [19]. Data of two hundred sixty-eight children were utilized in this research. The automated model performed well with an accuracy of 98.4%. Baik et al. [20] used a boosting algorithm to classify COVID-19 and non-COVID-19 illnesses. The xgboost classifier obtained a maximum AUC of 0.99. Shapley additive values (SHAP) was the XAI technique used, and the most critical markers according to it were D-Dimer, glucose, basophil and eosinophils. Another research used laboratory parameters to detect COVID-19 [21]. Five hundred fifteen patients from Poland were considered. The various models obtained accuracies between 90%-100%. Gomes et al. [22] used haematological markers for COVID-19 diagnosis. Multiple ML models were used, and an accuracy, specificity and sensitivity of 99% were obtained. According to the COVID-19 research, the most important features were troponin, ferritin, serum glucose, indirect bilirubin, D-dimer and lactate dehydrogenase (LDH).

The above studies show that COVID-19 can be diagnosed using machine learning and routine blood markers. However, no other studies exist that using machine learning to diagnose mild COVID-19 from similar influenzas. Our retrospective study also makes other significant contributions which are listed below:

- Descriptive and inferential statistical analysis has been utilized to comprehend the data better. Several exciting trends were observed using visualization techniques such as box plots and scatter plots.
- Ten nature-inspired techniques were utilized to select the important clinical attributes and were also compared further.
- Multiple mac The XAI should be used to explain a dataset from the data used with full explanation to a lay manhines learning algorithms, including stacking and voting, have been utilized. Two deep learning classifiers were also tested. The results obtained by the neural networks were compared with machine learning classifiers.
- The classification results attained by the best-performing algorithms were demystified utilizing five XAI approaches.
- The results obtained by the classifiers and the marker variations were also analyzed in-depth from a medical frame of reference.

The following sections are organized as follows: Materials and methods are detailed in Section II. Details about the dataset, statistical analysis, pre-processing, feature selection

TABLE 1. Markers utilized in research to predict COVID-19 diagnosis.

Number	Biomarker	Description	Number	Biomarker	Description
1	Age (In years)	Patient's age. Only patients above eighteen years of age were considered	14	Creatinine (mg/dL)	Creatinine is a chemical waste product of creatine.
2	Gender	Sex of a patient	15	Sodium (mmol/L)	Sodium content in the body.
3	Hemoglobin (gram/dL)	It is part of red blood cells and transfers oxygen.	16	Potassium (mmol/L)	Potassium content in the body.
4	Hematocrit (%)	It is a ratio of percentage by volume of red blood cells	17	Total. Bilirubin (T. Bilirubin) (mg/dL)	Liver enzyme.
5	TWBC (Total white blood cell count) (10^3/microliter)	TWBC contains neutrophils, lymphocytes, monocytes, eosinophils and basophils.	18	Direct Bilirubin (D. Bilirubin) (mg/dL)	Liver enzyme.
6	Neutrophil (%)	Part of white blood cells.	19	AST (IU/L)	Liver enzyme.
7	Lymphocyte (%)	Part of white blood cells.	20	ALT (IU/L)	Liver enzyme.
8	NLR (Neutrophil to Lymphocyte ratio)	Total number of neutrophils by total number of lymphocytes.	21	ALP (IU/L)	Liver enzyme.
9	Monocyte (%)	Part of white blood cells.	22	Protein (g/dL)	Protein content in the body.
10	Eosinophil (%)	Part of white blood cells.	23	Albumin (g/dL)	It is the most abundant circulating protein found in plasma
11	Platelet (%)	It helps in forming blood clots.	24	HbA1c (%)	A common test used to diagnose type 2 diabetes
12	Basophil (%)	Part of white blood cells.	25	Label (COVID-19/non-COVID-19)	Indicates whether the person has COVID-19 or non-COVID-19 infection
13	Urea (mg/dL)	Urea is the primary metabolite derived from dietary protein and tissue protein			

and machine learning methods have been included. The results and XAI interpretations are explained in-depth in Section III. Predictions and interpretations are discussed in-depth in Section IV. Section V concludes the article.

II. MATERIALS AND METHODS

A. DATASET FROM TWO HOSPITALS IN INDIA

Blood test data was attained from patients admitted to ''Dr TMA Pai Hospital'' and ''Kasturba Medical College''. Ethical clearance was procured before conducting this retrospective study (IEC: 613/2021). Patient information from RT-PCR tests performed between April 2022 and December 2022 was considered. The dataset was collected manually from files and Electronic Health Records (EHR). Only mild COVID-19 patients were considered for this research. The cohort consisted of 309 mild COVID-19 patients and 135 patients who exhibited similar symptoms but were considered COVID-19 negative by the RT-PCR test. The number of demographic and clinical attributes included was 25, which included the class label (RT-PCR test results). The attributes considered, their descriptions and units, are listed in Table 1.

B. STATISTICAL ANALYSIS AND DATA PREPROCESSING

Descriptive and inferential statistical analysis was made in this study using Jamovi. Jamovi is being used by data scientists and biostatisticians worldwide [23]. In the beginning, we have tabulated descriptive statistical measures for each attribute. Further, to gain inferences from the data, various t tests have been utilized to compare the two groups. Comparison of the two groups were also done using violin and scatter plots. Descriptive measures such as mean, median, sample size, minimum and maximum value and other measures are described in Table 2. All the above statistical measures are obtained for both COVID-19 and non-COVID-19 cohorts. To conduct t-tests, three variants of it have been utilized. They are the Student's t-test, Man-Whitney U test and the Welch's t-test. The marker is considered significant in differentiating between the cohorts if the obtained p-value is less than 0.001. The p-value obtained by the t-tests are detailed in Table 3. According to the above t-tests, the most critical markers were hemoglobin, TWBC, neutrophils, lymphocytes, NLR, monocytes, eosinophils, platelets, sodium, T. bilirubin, D. bilirubin, albumin and ALP.

Further, the difference in marker levels was analyzed using violin plots, as shown in Fig. 2. It can be observed that haemoglobin, lymphocytes and monocytes were higher in COVID-19 cases. Markers such as age, TWBC, neutrophil, eosinophil and platelets were lesser in the COVID-19 cohort. Scatter plots were then used to understand the relationships between markers, as depicted in Fig. 3. A direct relationship exists among haemoglobin and hematocrit. Albumin and protein were also linearly separable. A slight dependent relationship exists between platelets and TWBC. Further, when the neutrophil count increases, the lymphocyte count decreases (linear relationship).

The following steps were carried out during data pre-processing. Null values were removed, categorical variables were encoded, the data was balanced and the data was scaled. Generally, null values are substituted by measurements like mean and median. In this study, the median was preferred

TABLE 2. Descriptive statistical measures obtained for this study (COVID-19 and non-COVID-19 cohort).

	Cohort	Sample size chosen	Mean value obtained	Median value obtained	Standard deviation obtained	Inter quartile range obtained	Minimum value obtained	Maximum value obtained
Age	Non-COVID-19	135	52.711	54	19.9663	32.5	18	100
	COVID-19	309	47.32	48	19.4373	30	10	99
Hb(Hemoglobin)	Non-COVID-19	135	12.305	12.4	1.8276	1.65	6.1	17.3
	COVID-19	309	13.15	13.1	1.9066	2.3	3.7	17.4
PCV%(Haemtocrit)	Non-COVID-19	135	36.406	36.5	5.1427	4.65	19.45	51.2
	COVID-19	309	40.756	39	24.5768	7	9	365
TWBC	Non-COVID-19	135	8.497	7.95	4.3239	2.35	1.2	35.1
	COVID-19	309	5.587	5.3	2.0146	2.2	0.8	14.1
Neutrophil	Non-COVID-19	135	66.599	67.2	12.8779	14.35	28.1	93.7
	COVID-19	309	61.563	62	11.5501	15	25	95
Lymphocyte	Non-COVID-19	135	22.342	21.8	11.1797	13.55	2.5	56
	COVID-19	309	27.331	26	11.0344	13	2.1	91
NLR	Non-COVID-19	135	4.615	3	5.41	2.5	1	41
	COVID-19	309	2.534	2	2.9896	2	0	34
Monocyte	Non-COVID-19	135	8.223	7.8	3.2238	3.05	1	27.6
	COVID-19	309	9.602	9	3.6966	5	0	21
Eosinophil	Non-COVID-19	135	1.99	1.5	1.9079	1.65	0	8.3
	COVID-19	309	1.157	0.5	1.8009	1.3	0	13.9
Platelet	Non-COVID-19	135	266.642	263.5	111.9686	107.5	0	679
	COVID-19	309	233.559	222	88.3988	96	2.04	602
Basophil	Non-COVID-19	135	0.492	0.4	0.4234	0.3	0	2.5
	COVID-19	309	0.728	0.4	5.2202	0.2	0	92
Urea	Non-COVID-19	135	28.974	21.5	38.371	5.5	8	419
	COVID-19	309	23.152	21	13.1015	10	6	118
Creatinine	Non-COVID-19	135	0.938	0.8	0.6827	0.2	0.4	7.7
	COVID-19	309	0.874	0.8	0.3435	0.3	0.3	4.7
Sodium	Non-COVID-19	135	133.911	135	5.1724	3.5	112	142
	COVID-19	309	136.852	138	8.9427	5	1.28	148
Potassium	Non-COVID-19	135	4.126	4.1	0.3878	0.3	3.2	6
	COVID-19	309	4.238	4.1	2.0938	0.6	2.9	40
T. Bilirubin	Non-COVID-19	135	0.716	0.5	1.1288	0	0.2	12.4

TABLE 2. *(Continued.)* **Descriptive statistical measures obtained for this study (COVID-19 and non-COVID-19 cohort).**

	COVID-19	309	0.467	0.4	0.2426	0.3	0	1.6
D.Bilirubin	Non-COVID-19	135	0.362	0.2	0.9008	0	0.1	9.7
	COVID-19	309	0.191	0.2	0.0985	0.1	0.07	0.7
AST	Non-COVID-19	135	46.719	33	63.1078	0	10	599
	COVID-19	309	35.217	29	23.0595	20	0.2	229
ALT	Non-COVID-19	135	41.648	35	34.9737	1.75	9	266
	COVID-19	309	32.291	25	23.9823	22	5	180
ALP	Non-COVID-19	135	95.622	89	44.2149	0	35	504
	COVID-19	309	79.456	77	31.1166	28	17	382
Protein	Non-COVID-19	135	7.021	7	0.4147	0	5.9	9
	COVID-19	309	7.163	7.2	0.4714	0.1	4.2	8.5
Albumin	Non-COVID-19	135	3.847	3.9	0.3492	0	1.5	4.6
	COVID-19	309	4.267	4.3	0.3869	0.1	3	7
HbA1c	Non-COVID-19	135	6.1	5.8	1.3131	0	4	13.1
	COVID-19	309	6.12	5.6	1.5407	1	4	14

since they deal better with outliers [24]. Categorical encoding is essential in preprocessing [25]. Gender was the only categorical attribute available in this data. One-hot encoding method was deployed to encode the gender attribute. The above technique was used since it is generally reliable [26]. The data needed to be more balanced in research. The number of mild COVID-19 patients considered was 309. The non-COVID-19 cohort consisted of only 135 patients. When cohorts are imbalanced, results are biased [27]. The models also consider the minor class as outliers. The borderline synthetic minority oversampling technique (SMOTE) was deployed to balance the data [28]. The oversampling technique is a better version of the traditional SMOTE since it also considers the decision boundary while generating synthetic data. The test dataset was not exposed to data balancing to protect the dataset's integrity. The standardization technique was utilized for scaling the data [29]. This technique converts all the values between '-1' and '1' using standard deviation. It is crucial to scale the data to prevent the classifiers from preferring higher values over lower ones.

C. FEATURE SELECTION

The performance of the ML classifiers reduces when unnecessary features are chosen for model training. The main aim of feature selection is to identify the optimal set of variables that will allow the building of best performing models [30]. Generally, three types of methods are used: (a) Filter methods, (b) Wrapper methods (c) Embedded methods [31]. Filter methods use univariate statistics to choose the most essential features. They are also faster in execution. Wrappers need an approach for searching the available space of all potential characteristic subsets, evaluating each feature's accuracy level by building and assessing a classifier using it. The wrapper algorithms are more efficient than filter methods. However, the filter methods are faster. The embedded method combines the above two techniques. In this study, two filter and ten wrapper methods were utilized.

Pearson's correlation method describes the degree of association among two parameters [32]. The Pearson's correlation coefficient exists between -1 and 1. When the number is closer to '1', it indicates a positive correlation.

The number near '-1' indicates that the variables are negatively correlated. It has several strengths. It is easy to calculate and understand. It is a standardized measure which calculates the strength and direction of relationships. It can also be utilized for bivariate analysis. However, it is sensitive to outliers. It also assumes linearity and does not capture causation. The Pearson's heat map is described in Fig. 4. According to the figure, the important markers are TWBC, NLR and Albumin. The top 10 features were chosen using this methodology, which are tabulated in Table 4. Mutual information (MI) has also been used in this study. It uses the concepts of ''mutual dependence'', ''information theory'', and ''probability theory'' [33]. The entropy of the feature is also measured in this algorithm. Advantages of the algorithm

TABLE 3. Inferential statistics conducted using three t tests (a) Student's t test (S) (b) Welch's t test (W) (c) Mann-Whitney U test (M). A marker is considered important if p-value is less than <.001.

Marker	t test	p-value	Marker	t test	p-value
Age	S	0.008	Creatinine	S	0.186
	W	0.009		W	0.299
	M	0.009		M	0.745
Hb(Hemoglobin)	S	< .001	Sodium	S	< .001
	W	< .001		W	< .001
	M	< .001		M	< .001
PCV%(Haemtocrit)	S	0.042	Potassium	S	0.538
	W	0.003		W	0.366
	M	< .001		M	0.983
TWBC	S	< .001	T. Bilirubin	S	< .001
	W	< .001		W	0.013
	M	< .001		M	< .001
Neutrophil	S	< .001	D.Bilirubin	S	0.001
	W	< .001		W	0.03
	M	< .001		M	< .001
Lymphocyte	S	< .001	AST	S	0.005
	W	< .001		W	0.041
	M	< .001		M	0.002
NLR	S	< .001	ALT	S	0.001
	W	< .001		W	0.005
	M	< .001		M	< .001
Monocyte	S	< .001	ALP	S	< .001
	W	< .001		W	< .001
	M	< .001		M	< .001
Eosinophil	S	< .001	Protein	S	0.003
	W	< .001		W	0.002
	M	< .001		M	< .001
Platelet	S	< .001	Albumin	S	< .001
	W	0.003		W	< .001
	M	< .001		M	< .001
Basophil	S	0.6	HbA1c	S	0.897
	W	0.43		W	0.89
	M	0.683		M	0.002
Urea	S	0.018			
	W	0.088			
	M	0.236			

include capturing non-linear relationships, robustness and faster execution. It also does not make random assumptions about the data. However, there are a few disadvantages too. The algorithm is less efficient when high dimensional data is used. It is also sensitive to bias. Interpretation of the algorithm can also be a challenge. The most critical biomarkers are described in Fig. 5. The essential markers are Albumin, Protein, HbA1c, AST and ALT. The top ten features were chosen and are described in Table 4.

Further, ten nature-inspired algorithms have also been utilized. These methods are influenced by natural processes which occur in everyday life [34]. They support parallel and distributed computing and focus on global optimization. They are also robust and versatile. However, computational cost can be an issue. It is also prone to parameter sensitivity. Sometimes, it is also difficult to understand them theoretically. The markers chosen by these algorithms are also tabulated in Table 4.

The clinical attributes selected by various feature selection algorithms are depicted in Figure 6. From the diagram, it can be seen that ALP, ALT, haemoglobin and TWBC were the most chosen markers. Many algorithms also chose sodium and eosinophils. However, platelets were not considered by any technique.

D. MACHINE LEARNING METHODOLOGY

This study considered several ML techniques, such as bagging, boosting, stacking and voting. Hyperparameter optimization was conducted using the grid search method. This searching technique finds the optimal hyperparameters for each algorithm. It is a simple technique and has already been used in several medical studies [35]. The five-fold cross-validation was utilised to shuffle the training and testing data into various sets. The training data is equally divided into five subsets, and the validation is done separately for each subset. The results are then averaged to make the models more reliable. Further, the stacking approach was used to combine eight baseline classifiers.

Stacking uses an ensemble technique to merge the predictions made by the multiple models [36]. Hence, the stacked models become more reliable than an individual model. Stacking also makes use of a ''meta-learner'', which helps in making the predictions. Voting combines the predictions of multiple classifiers and selects the class using the ''majority vote'' [37]. In this study, we have considered hard-voting and soft-voting classifiers. In hard voting, each algorithm predicts a class using the concept of the highest vote share. In soft voting, each model generates a probability score. The average score of probabilities is chosen for predicting the class label.

Two deep learning algorithms: DNN (Deep Neural Network) and 1D-CNN (1-Dimensional Convolutional Neural Networks) were also tested to predict the diagnosis [38], [39], [40], [41]. Due to their capacity to independently predict and retrieve hierarchical characteristics from unprocessed information, DNNs have achieved exceptional success across various fields, including medicine. A CNN is typically used

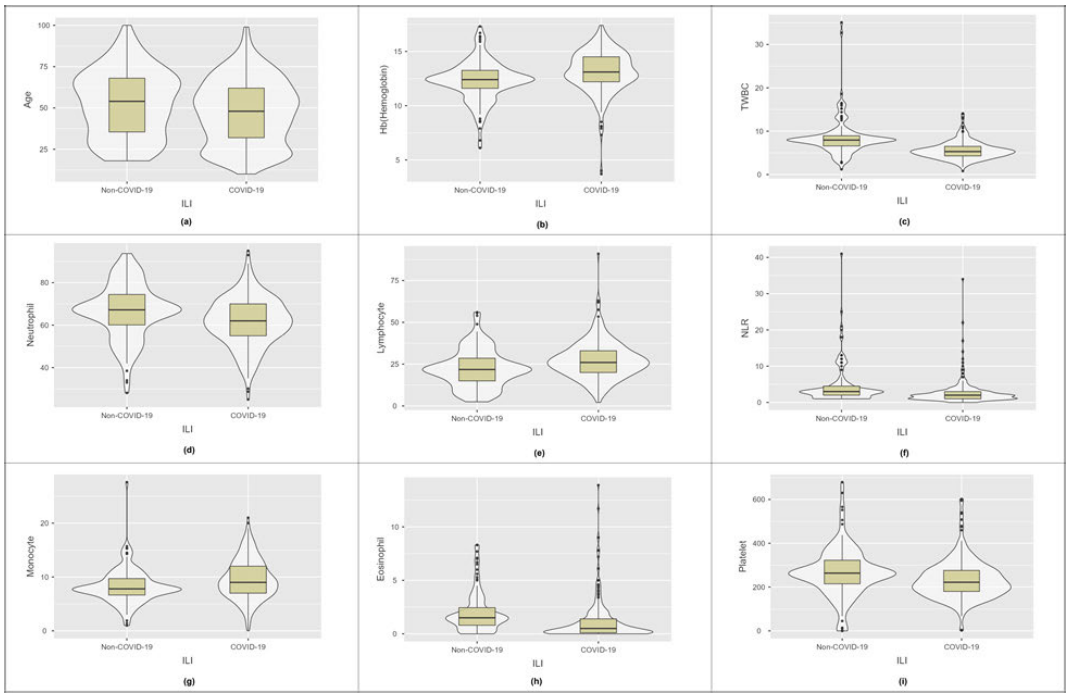

FIGURE 2. Violin plots for Non-COVID-19 and COVID-19 groups. (a) Patient age (b) Hemoglobin (c)TWBC (d) Neutrophil (e) Lymphocyte (f) NLR (g) Monocyte (h) Eosinophil (i) Platelet.

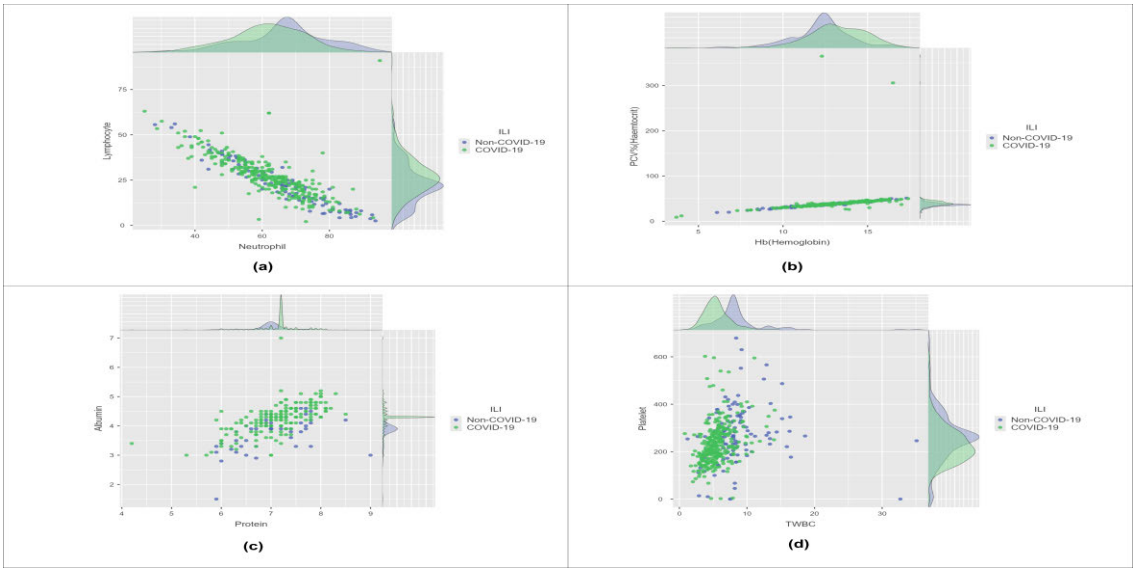

FIGURE 3. Scatter plots to understand the relationship between variables. (a) Lymphocyte vs Neutrophil (b) Hematocrit vs Hemoglobin (c) Albumin vs Protein (d) Platelet vs TWBC.

to predict image data. However, 1D-CNN is specifically designed for time series data analysis.

The test dataset was evaluated using various classification and loss metrics. The results obtained by the classifier were understood using five heterogeneous explainers. The importance of each feature in predicting the model output can be analyzed accurately using these algorithms. The entire process carried out in this research is clearly depicted in Figure 7. In the beginning, data was collected from two hospitals. Descriptive and inferential statistical analysis was conducted on the data later. Various preprocessing methods such as removal of empty values, categorical variable encoding and scaling were done later. Twelve feature selection methodologies were further utilized to select the best markers. After data splitting, the training data was then balanced using Borderline SMOTE. Further, machine

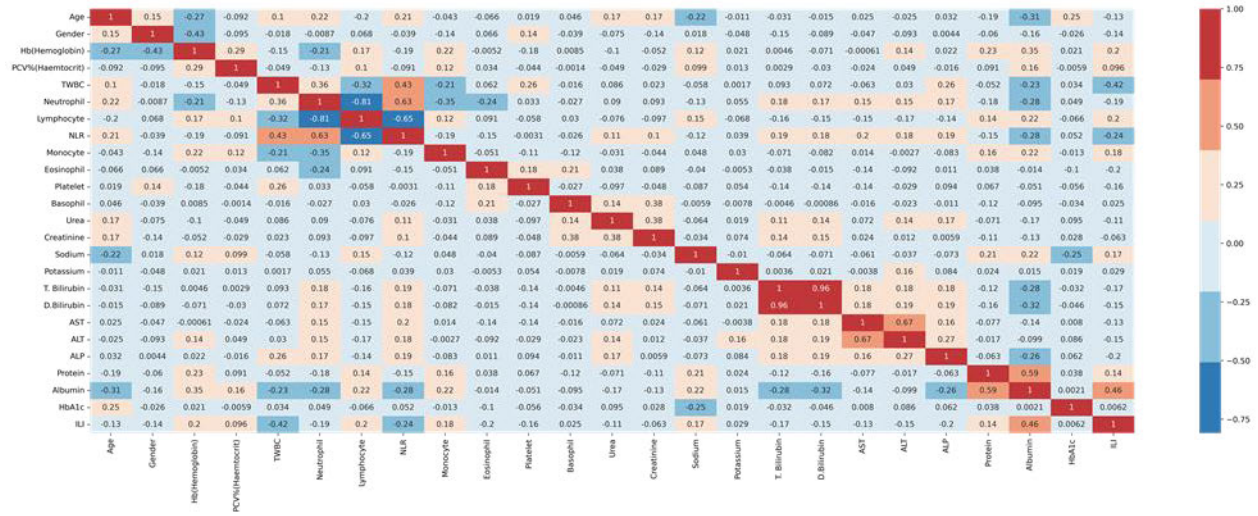

FIGURE 4. Finding important biomarkers for COVID-19 detection using Pearson's correlation heatmap.

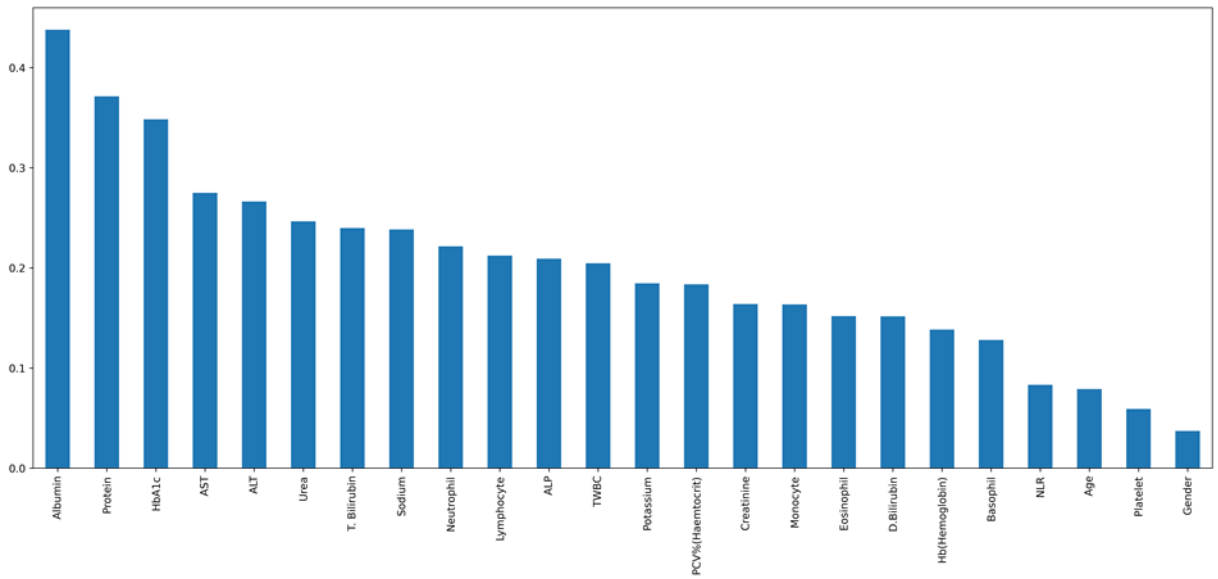

FIGURE 5. Feature selection using mutual importance.

learning models were trained and tested. The predictions were also compared with deep learning models. XAI was then used to interpret the classifiers. The models could then be deployed to perform real-time diagnosis on new prospective data.

III. RESULTS

A. MODEL EVALUATION

Multiple ML models were tested in this study. Every classifier was trained on ten occasions, with the average being calculated. This procedure was performed to prevent overfitting. The weighted average F1 scores attained by the classifiers are tabulated in Table 5. Among all the algorithms, the optimal results were obtained by the mutual information, Pearson's correlation, bat algorithm, flower pollination algorithm, particle swarm optimization and salp swarm optimization. The weighted F1-score obtained by the stacked model were 94%, 89%, 88%, 92%, 90% and 87%, respectively. The above feature selection methods were considered for further analysis since they obtained good results. The classification and loss metrics for the above-selected algorithms are described in Table 6. When the mutual information algorithms were used, the optimal predictions were obtained by the xgboost and catboost models with an outstanding accuracy of 95%. The combined stacking, hard voting (HV) and soft voting (SV) classifiers attained accuracies of 94%, 90% and 90%, respectively. When Pearson's correlation was utilized, an accuracy of 93% was

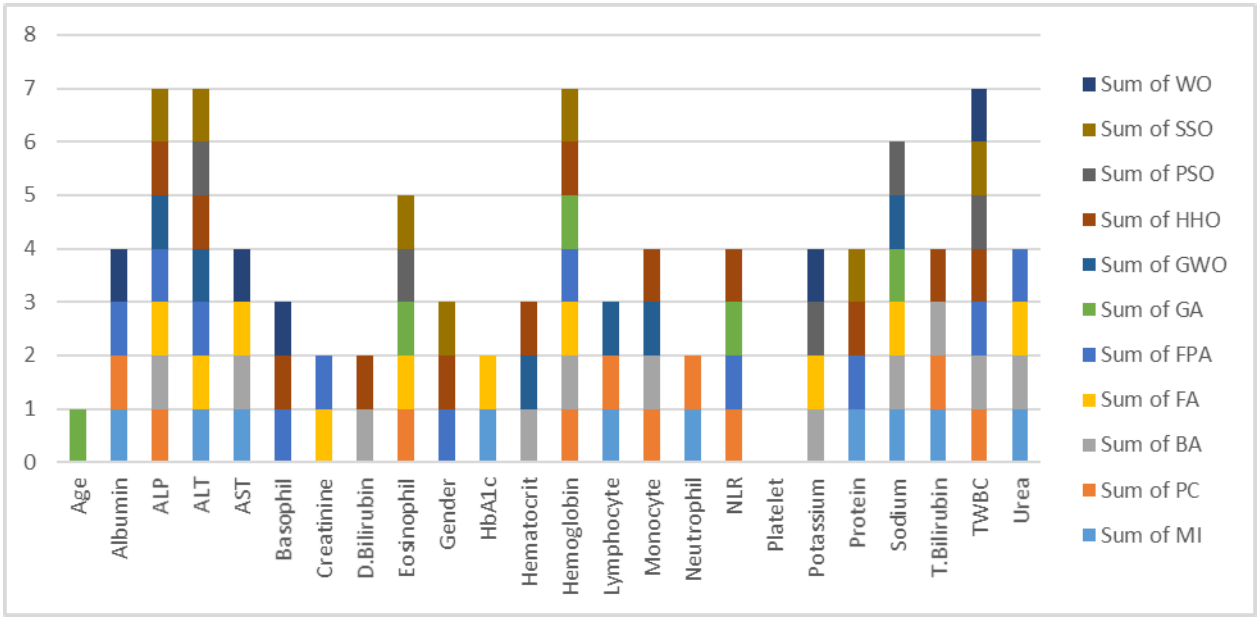

FIGURE 6. Markers chosen by the feature selection techniques.

TABLE 4. The number of clinical attributes chosen by each technique along with the feature names.

Feature selection Technique utilized	Number of biomarkers chosen	Biomarker names
Mutual information (MI)	10	Albumin, protein, HbA1c, AST, ALT, Urea, Bilirubin, Sodium, Neutrophil and Lymphocyte.
Pearson's correlation (PC)	10	Albumin, TWBC, NLR, Eosinophil, ALP, Lymphocyte, Hemoglobin, Neutrophil, Monocyte and Bilirubin.
Nature inspired feature selection		
Bat algorithm (BA) [42]	11	Hematocrit, Hemoglobin, TWBC, Monocyte, Urea, Sodium, Potassium, T. Bilirubin, D. Bilirubin, AST and ALP.
Cuckoo search algorithm (CSA) [43]	7	Eosinophil, Hemoglobin, TWBC, Basophil, AST and ALT.
Firefly algorithm (FA) [44]	10	Eosinophil, Hemoglobin, Sodium, Potassium, AST, Urea, Creatinine, ALT, ALP and HbA1c.
Flower pollination algorithm (FPA) [45]	11	Gender, Hemoglobin, TWBC, NLR, Basophil, Urea, Creatinine, ALT, ALP, Protein and Albumin.
Genetic algorithm (GA) [46]	5	Age, Hemoglobin, NLR, Eosinophil and Sodium.
Grey wolf optimizer (GWO) [47]	6	Hematocrit, Lymphocyte, Sodium, Monocyte, ALT and ALP.
Harris hawk's optimizer (HHO) [48]	12	Gender, Hemoglobin, Hematocrit, TWBC, NLR, T. Bilirubin, Bilirubin, ALT, ALP, Protein Monocyte and Basophil, .
Particle swarm optimization (PSO) [49]	5	TWBC, Eosinophil, Sodium, Potassium and ALT.
Salp swarm optimization (SSO) [50]	7	Gender, Hemoglobin, Protein, Eosinophil, TWBC, ALT and ALP.
Whale optimization (WO) [51]	5	Albumin, Potassium, Basophil, AST and TWBC.

attained by the catboost algorithm. The stacking, HV and SV attained accuracies of 89%, 91% and 98% each. When the bat algorithm was used, the maximum accuracy was obtained by the random forest model (90%). The stacking, HV and SV algorithms attained accuracies of 88%, 84% and 83%, respectively. When the flower pollination method was utilized, an accuracy of 94% was attained by the xgboost classifier. The stacking, HV and SV algorithms attained accuracies of 92%, 90% and 90%, respectively. When the particle swam optimization was utilized, an accuracy of 90% was attained by the lightgbm and stacking classifiers. HV and SV attained accuracies of 82% and 86%, respectively. When the salp swarm optimization was utilized, an accuracy of 93% was attained by the lightgbm classifier. The ensemble classifiers (stacking, HV and SV) attained accuracies of 87% each. The hyperparameters chosen by the algorithms are described in Table 7. The AUC curve graphs for the stacking algorithms are depicted in Fig. 8. An AUC of 0.97 was attained by particle swarm optimization and mutual information methods. From the figure, it can also

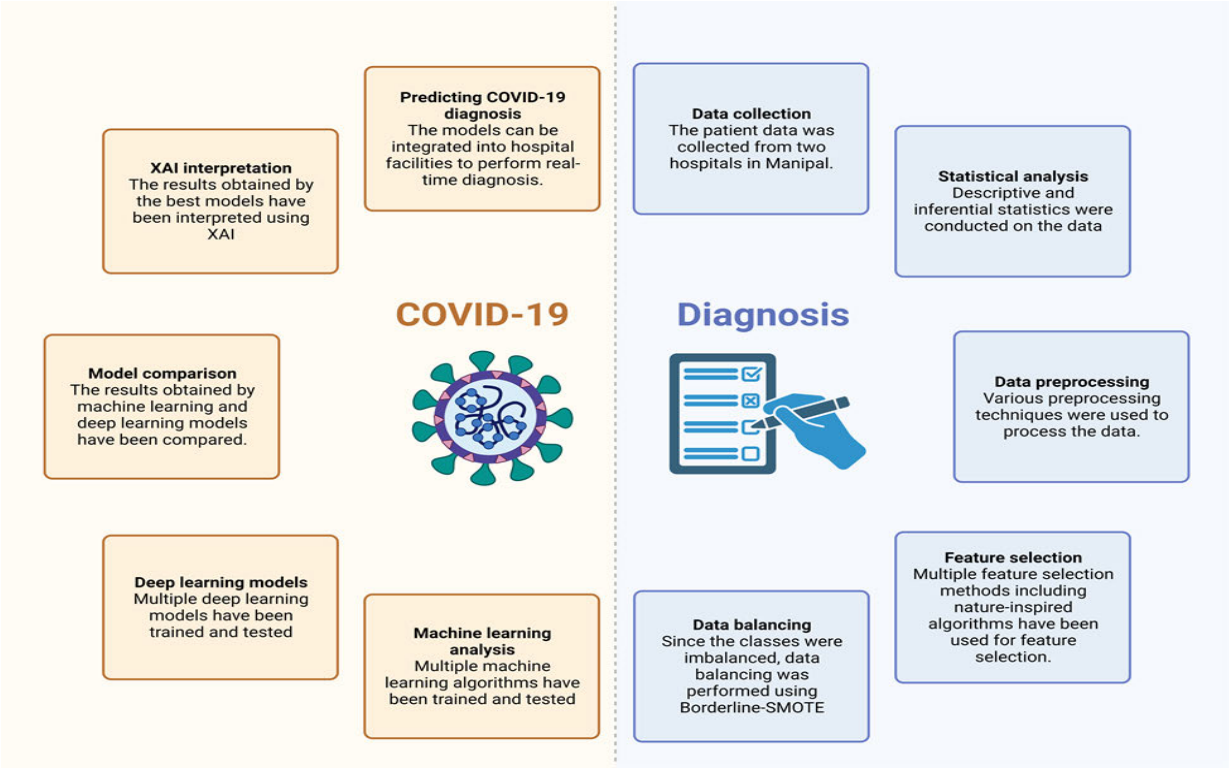

FIGURE 7. Process flow of this research.

TABLE 5. Weighted average F1-score obtained in this research.

Algorithm	Weighted average F1-score											
	MI	PC	BA	CSA	FA	FPA	GA	GWO	HHO	PSO	SSO	WO
Random forest	93%	89%	90%	80%	86%	92%	86%	77%	87%	89%	89%	86%
Logistic regression	86%	83%	75%	76%	78%	85%	76%	74%	83%	86%	79%	82%
Decision tree	88%	82%	79%	74%	83%	86%	72%	66%	81%	87%	77%	80%
KNN	80%	80%	75%	82%	74%	84%	73%	69%	75%	85%	72%	86%
Adaboost	91%	89%	85%	84%	82%	89%	76%	77%	89%	89%	90%	82%
Catboost	95%	92%	89%	83%	83%	92%	85%	76%	88%	87%	89%	83%
Lightgbm	93%	89%	88%	83%	86%	93%	86%	78%	90%	90%	93%	85%
Xgboost	95%	90%	89%	82%	87%	94%	82%	83%	89%	89%	92%	84%
Stacking	94%	89%	88%	85%	86%	92%	77%	79%	86%	90%	87%	82%
Hard voting	90%	91%	84%	87%	89%	89%	83%	82%	88%	87%	87%	87%
Soft voting	90%	90%	83%	87%	88%	90%	83%	81%	86%	86%	88%	83%

be inferred that the number of false negative and false positive cases were very few. The precision-recall curves for the optimal six feature selection algorithms (stacked algorithms) are pictorially depicted in Fig 9. The highest maximum precision was obtained for the mutual information method. Most algorithms obtained good precision and recall values according to the figure. The confusion matrices for the optimal six feature selection algorithms (stacked classifiers) are described in Fig. 10. From the figure, it can be inferred that the number of wrongly classified cases were minimal for the mutual information and flower pollination algorithm.

Further, four deep neural networks are utilized. The architecture for DNN, 1D-CNN and LSTM have been depicted in Fig 11. The results attained by the algorithms are depicted in Table 8. The 1D-CNN attained an accuracy of 83% among all the classifiers. Compared to machine learning algorithms, the DL models' results were inferior. This is a common trend in medical AI research when the data is small [38], [39].

TABLE 6. Classification metrics obtained in this research.

Mutual Information									
Algorithm	Accuracy	Precision	Recall	AUC	Average precision	Log loss	Hamming loss	Jaccard score	Mathew's correlation coefficient
Random forest	93%	93%	93%	0.98	0.99	2.57	0.07	0.90	0.80
Logistic regression	86%	88%	86%	0.89	0.96	4.38	0.12	0.83	0.68
Decision tree	87%	88%	87%	0.86	0.92	4.21	0.10	0.85	0.69
KNN	79%	81%	79%	0.76	0.87	7.21	0.20	0.74	0.49
Adaboost	91%	92%	91%	0.97	0.99	3.09	0.08	0.88	0.78
Catboost	95%	95%	95%	0.99	0.99	1.81	0.05	0.93	0.86
Lightgbm	93%	93%	93%	0.97	0.99	2.31	0.06	0.91	0.824
Xgboost	95%	95%	95%	0.99	0.99	1.80	0.05	0.92	0.86
Stacking	94%	94%	94%	0.97	0.99	2.06	0.05	0.92	0.84
Hard voting	90%	90%	90%	0.95	0.92	3.60	0.10	0.85	0.77
Soft Voting	90%	91%	90%	0.95	0.98	3.35	0.09	0.86	0.77
Pearson's Correlation									
Algorithm	Accuracy	Precision	Recall	AUC	Average precision	Log loss	Hamming loss	Jaccard score	Mathew's correlation coefficient
Random forest	89%	89%	89%	0.95	0.97	3.86	0.11	0.85	0.72
Logistic regression	83%	85%	83%	0.86	0.94	5.92	0.17	0.77	0.61
Decision tree	81%	83%	81%	0.86	0.91	6.44	0.18	0.75	0.58
KNN	80%	79%	80%	0.74	0.83	6.95	0.20	0.75	0.50
Adaboost	89%	89%	89%	0.96	0.98	3.86	0.11	0.85	0.73
Catboost	93%	92%	93%	0.94	0.96	2.57	0.07	0.90	0.81
Lightgbm	90%	89%	90%	0.92	0.96	3.60	0.10	0.86	0.74
Xgboost	90%	90%	90%	0.95	0.97	3.35	0.09	0.87	0.75
Stacking	89%	89%	89%	0.93	0.97	3.86	0.11	0.85	0.72
Hard voting	91%	91%	91%	0.96	0.92	3.09	0.08	0.87	0.80
Soft Voting	90%	90%	90%	0.95	0.96	3.60	0.10	0.85	0.76
Bat Algorithm									
Algorithm	Accuracy	Precision	Recall	AUC	Average precision	Log loss	Hamming loss	Jaccard score	Mathew's correlation coefficient
Random forest	90%	91%	90%	0.95	0.96	3.35	0.09	0.85	0.79
Logistic regression	75%	80%55	75%	0.82	0.9	8.76	0.25	0.63	0.52
Decision tree	78%	80%	78%	0.84	0.87	7.47	0.21	0.69	0.55
KNN	75%	76%	75%	0.75	0.79	8.76	0.25	0.65	0.47
Adaboost	85%	87%	85%	0.9	0.93	5.15	0.14	0.78	0.70
Catboost	89%	89%	89%	0.95	0.96	3.86	0.11	0.83	0.76
Lightgbm	88%	89%	88%	0.95	0.97	4.12	0.11	0.82	0.75
Xgboost	89%	89%	89%	0.95	0.96	3.86	0.11	0.83	0.76
Stacking	88%	88%	88%	0.95	0.97	4.12	0.11	0.82	0.75
Hard voting	84%	84%	84%	0.09	0.05	5.67	0.16	0.78	0.63
Soft Voting	83%	83%	83%	0.89	0.93	5.12	0.17	0.77	0.60
Flower Pollination Algorithm									
Algorithm	Accuracy	Precision	Recall	AUC	Average precision	Log loss	Hamming loss	Jaccard score	Mathew's correlation coefficient
Random forest	92%	92%	92%	0.97	0.99	2.83	0.08	0.89	0.80
Logistic regression	85%	86%	85%	0.91	0.96	5.15	0.14	0.80	0.64
Decision tree	86%	87%	86%	0.85	0.91	4.89	0.14	0.81	0.66
KNN	84%	84%	84%	0.81	0.88	5.42	0.15	0.80	0.61
Adaboost	90%	89%	90%	0.91	0.94	3.60	0.10	0.86	0.73
Catboost	92%	92%	92%	0.98	0.99	2.83	0.08	0.89	0.80
Lightgbm	93%	93%	93%	0.97	0.99	2.57	0.07	0.90	0.81
Xgboost	94%	94%	94%	0.97	0.99	2.06	0.05	0.92	0.92
Stacking	92%	92%	92%	0.96	0.98	2.03	0.08	0.89	0.79
Hard voting	90%	90%	90%	0.94	0.91	3.86	0.11	0.85	0.73
Soft Voting	90%	90%	90%	0.92	0.96	3.60	0.10	0.86	0.75
Particle Swarm Optimization									

TABLE 6. *(Continued.)* **Classification metrics obtained in this research.**

Algorithm	Accuracy	Precision	Recall	AUC	Average precision	Log loss	Hamming loss	Jaccard score	Mathew's correlation coefficient
Random forest	89%	89%	89%	0.96	0.98	3.80	0.11	0.84	0.74
Logistic regression	86%	88%	86%	0.91	0.95	4.89	0.10	0.80	0.70
Decision tree	87%	88%	87%	0.94	0.96	4.63	0.13	0.81	0.71
KNN	85%	86%	85%	0.84	0.88	5.15	0.14	0.79	0.66
Adaboost	89%	89%	89%	0.95	0.97	3.86	0.11	0.84	0.75
Catboost	87%	87%	87%	0.95	0.98	4.63	0.13	0.81	0.69
Lightgbm	90%	90%	90%	0.97	0.99	3.60	0.10	0.85	0.76
Xgboost	89%	89%	89%	0.97	0.98	3.86	0.11	0.84	0.75
Stacking	90%	91%	90%	0.97	0.98	3.35	0.18	0.86	0.78
Hard voting	87%	87%	87%	0.93	0.88	4.63	0.13	0.82	4.63
Soft Voting	86%	86%	86%	0.93	0.97	4.89	0.14	0.81	0.66
Salp Swarm Optimization									
Algorithm	Accuracy	Precision	Recall	AUC	Average precision	Log loss	Hamming loss	Jaccard score	Mathew's correlation coefficient
Random forest	90%	90%	90%	0.95	0.97	3.68	0.10	0.85	0.76
Logistic regression	79%	80%	79%	0.83	0.88	7.21	0.20	0.21	0.51
Decision tree	77%	78%	77%	0.85	0.68	7.90	0.23	0.68	0.512
KNN	72%	72%	72%	0.69	0.76	9.79	0.28	0.63	0.38
Adaboost	90%	90%	90%	0.94	0.97	3.35	0.69	0.86	0.78
Catboost	89%	89%	89%	0.94	0.97	3.86	0.11	0.84	0.75
Lightgbm	93%	93%	93%	0.95	0.97	2.57	0.07	0.89	0.83
Xgboost	92%	92%	92%	0.95	0.98	2.83	0.08	0.88	0.81
Stacking	87%	87%	87%	0.93	0.96	4.38	0.12	0.82	0.71
Hard voting	87%	87%	87%	0.94	0.95	4.63	0.13	0.82	0.69
Soft Voting	87%	88%	87%	0.94	0.97	4.38	0.12	0.82	0.71

B. EXPLAINABLE ARTIFICIAL INTELLIGENCE

This retrospective study uses five XAI techniques to understand the most critical markers in COVID-19 diagnosis. The explainers also make the models interpretable and understandable [10], [11]. The STACK models of the best six techniques were chosen for explanation. The STACK models performed well, and their predictions are trustworthy. Deep learning algorithms were not selected for analysis since their performance was slightly inferior to the machine learning algorithms. Further, many XAI techniques still need to work for deep learning models.

SHAP uses a probabilistic approach and a set of strategies for describing the outcomes of the classifiers [52]. It is based on assigning a Shapley value for every attribute and measuring its contribution. It also uses the concept of ''game theory''. The beeswarm plots obtained for all six feature selection techniques (stacking classifier) are depicted in Fig. 12. The hyperplane separates the two classes. The patients predicted as COVID-19 negative are placed towards the left of the hyperplane, and those diagnosed as COVID-19 positive are placed towards the right. Red depicts an elevated value of the biomarker, and blue depicts a lower value of the biomarker. The graphs show that the critical markers are Albumin, Protein, HbA1c, Sodium, Eosinophil, TWBC and ALP.

Albumin, Protein, and Sodium levels increased for the COVID-19 cohort, and ALP, HbA1c, Eosinophil, and TWBC decreased for the COVID-19 cohort.

LIME is a local explainer that helps explain the model predictions [53]. This algorithm's main steps are selecting an instance, perturbations, model predictions, and fitting the best model and interpreting it. LIME is very versatile and is used in various domains. The LIME graphs for the optimal six algorithms are described in Fig. 13. Fig. 13(a), Fig. 13(d), and Fig. 13(e) indicate a COVID-19 negative patient. Blood variables like Albumin, Protein, T. Bilirubin and TWBC point the prediction towards COVID-19 negative diagnosis. Fig. 13(b), Fig. 13(c) and Fig. 13(d) indicate a COVID-19 positive prediction. Laboratory biomarkers such as Albumin, Eosinophil, TWBC, Sodium and Protein have been essential for this diagnosis.

Eli5 is an explainer commonly used to decipher tree-based machine-learning models [54]. The Eli5 predictions are described in Fig. 14. The clinical parameters are arranged based on their contribution to the model output. The explainer also considers the bias (error) in the prediction. It can be seen that the most crucial attributes are Albumin, Protein, Lymphocytes, TWBC and Urea.

Qlattice uses ''quantum theory'' to explain the predictions [55]. It was developed by a company named

TABLE 7. Hyperparameters chosen by the machine learning algorithms.

Mutual Information	
Algorithm	Hyperparameters
Random forest	{'bootstrap': True, 'max_depth': 80, 'max_features': 2, 'min_samples_leaf': 3, 'min_samples_split': 8, 'n_estimators': 300}
Logistic regression	{'C': 1000, 'penalty': 'l2'}
Decision tree	{'criterion': 'entropy', 'max_depth': 150, 'max_features': 'log2', 'min_samples_leaf': 1, 'min_samples_split': 10, 'splitter': 'best'}
KNN	{'n_neighbors': 1}
Adaboost	{'learning_rate': 0.1, 'n_estimators': 200}
Catboost	{'border_count': 5, 'depth': 3, 'iterations': 250, 'l2_leaf_reg': 5, 'learning_rate': 0.03}
Lightgbm	{'lambda_l1': 1.5, 'lambda_l2': 0, 'min_data_in_leaf': 30, 'num_leaves': 31, 'reg_alpha': 0.1}
Xgboost	{'colsample_bytree': 0.3, 'gamma': 0.2, 'learning_rate': 0.15, 'max_depth': 3, 'min_child_weight': 3}
Pearson's Correlation	
Random forest	{'bootstrap': True, 'max_depth': 90, 'max_features': 2, 'min_samples_leaf': 3, 'min_samples_split': 8, 'n_estimators': 200}
Logistic regression	{'C': 100, 'penalty': 'l2'}
Decision tree	{'criterion': 'entropy', 'max_depth': 50, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 10, 'splitter': 'best'}
KNN	{'n_neighbors': 1}
Adaboost	{'learning_rate': 0.1, 'n_estimators': 300}
Catboost	{'border_count': 32, 'depth': 3, 'iterations': 250, 'l2_leaf_reg': 3, 'learning_rate': 0.03}
Lightgbm	{'lambda_l1': 0, 'lambda_l2': 0, 'min_data_in_leaf': 30, 'num_leaves': 31, 'reg_alpha': 0.1}
Xgboost	{'colsample_bytree': 0.4,

TABLE 7. *(Continued.)* **Hyperparameters chosen by the machine learning algorithms.**

	'gamma': 0.1, 'learning_rate': 0.05, 'max_depth': 3, 'min_child_weight': 1}
Bat Algorithm	
Random forest	{'bootstrap': True, 'max_depth': 100, 'max_features': 2, 'min_samples_leaf': 3, 'min_samples_split': 8, 'n_estimators': 300}
Logistic regression	{'C': 1000, 'penalty': 'l2'}
Decision tree	{'criterion': 'gini', 'max_depth': 9, 'max_features': 'auto', 'min_samples_leaf': 1, 'min_samples_split': 10, 'splitter': 'best'}
KNN	{'n_neighbors': 1}
Adaboost	{'learning_rate': 1.0, 'n_estimators': 200}
Catboost	{'border_count': 10, 'depth': 3, 'iterations': 250, 'l2_leaf_reg': 1, 'learning_rate': 0.03}
Lightgbm	{'lambda_l1': 0, 'lambda_l2': 1, 'min_data_in_leaf': 30, 'num_leaves': 31, 'reg_alpha': 0.1}
Xgboost	{'colsample_bytree': 0.3, 'gamma': 0.0, 'learning_rate': 0.15, 'max_depth': 6, 'min_child_weight': 1}
Flower Pollination Algorithm	
Random forest	{'bootstrap': True, 'max_depth': 100, 'max_features': 3, 'min_samples_leaf': 3, 'min_samples_split': 8, 'n_estimators': 1000}
Logistic regression	{'C': 100, 'penalty': 'l2'}
Decision tree	{'criterion': 'gini', 'max_depth': 5, 'max_features': 'log2', 'min_samples_leaf': 1, 'min_samples_split': 10, 'splitter': 'best'}
KNN	{'n_neighbors': 1}
Adaboost	{'learning_rate': 1.0, 'n_estimators': 100}
Catboost	{'border_count': 5, 'depth': 3, 'iterations': 250, 'l2_leaf_reg': 1, 'learning_rate': 0.03}
Lightgbm	{'lambda_l1': 0,

TABLE 7. *(Continued.)* **Hyperparameters chosen by the machine learning algorithms.**

	'lambda_l2': 0, 'min_data_in_leaf': 50, 'num_leaves': 31, 'reg_alpha': 0.1}
Xgboost	{'colsample_bytree': 0.3, 'gamma': 0.0, 'learning_rate': 0.15, 'max_depth': 4, 'min_child_weight': 1}
Particle Swarm Optimization	
Random forest	{'bootstrap': True, 'max_depth': 90, 'max_features': 2, 'min_samples_leaf': 3, 'min_samples_split': 8, 'n_estimators': 200}
Logistic regression	{'C': 1000, 'penalty': 'l2'}
Decision tree	{'criterion': 'entropy', 'max_depth': 12, 'max_features': 'log2', 'min_samples_leaf': 8, 'min_samples_split': 10, 'splitter': 'best'}
KNN	{'n_neighbors': 1}
Adaboost	{'learning_rate': 0.1, 'n_estimators': 300}
Catboost	{'border_count': 10, 'depth': 3, 'iterations': 250, 'l2_leaf_reg': 3, 'learning_rate': 0.03}
Lightgbm	{'lambda_l1': 0, 'lambda_l2': 0, 'min_data_in_leaf': 30, 'num_leaves': 31, 'reg_alpha': 0.1}
Xgboost	{'colsample_bytree': 0.4, 'gamma': 0.1, 'learning_rate': 0.1, 'max_depth': 3, 'min_child_weight': 1}
Salp Swarm Optimization	
Random forest	{'bootstrap': True, 'max_depth': 100, 'max_features': 2, 'min_samples_leaf': 3, 'min_samples_split': 8, 'n_estimators': 1000}
Logistic regression	{'C': 1000, 'penalty': 'l2'}
Decision tree	{'criterion': 'gini', 'max_depth': 11, 'max_features': 'auto', 'min_samples_leaf': 1, 'min_samples_split': 10, 'splitter': 'best'}
KNN	{'n_neighbors': 1}
Adaboost	{'learning_rate': 0.1, 'n_estimators': 1000}

TABLE 7. *(Continued.)* **Hyperparameters chosen by the machine learning algorithms.**

Catboost	{'border_count': 10, 'depth': 3, 'iterations': 250, 'l2_leaf_reg': 1, 'learning_rate': 0.03}
Lightgbm	{'lambda_l1': 1, 'lambda_l2': 0, 'min_data_in_leaf': 30, 'num_leaves': 31, 'reg_alpha': 0.1}
Xgboost	{'colsample_bytree': 0.3, 'gamma': 0.1, 'learning_rate': 0.1, 'max_depth': 8, 'min_child_weight': 1}

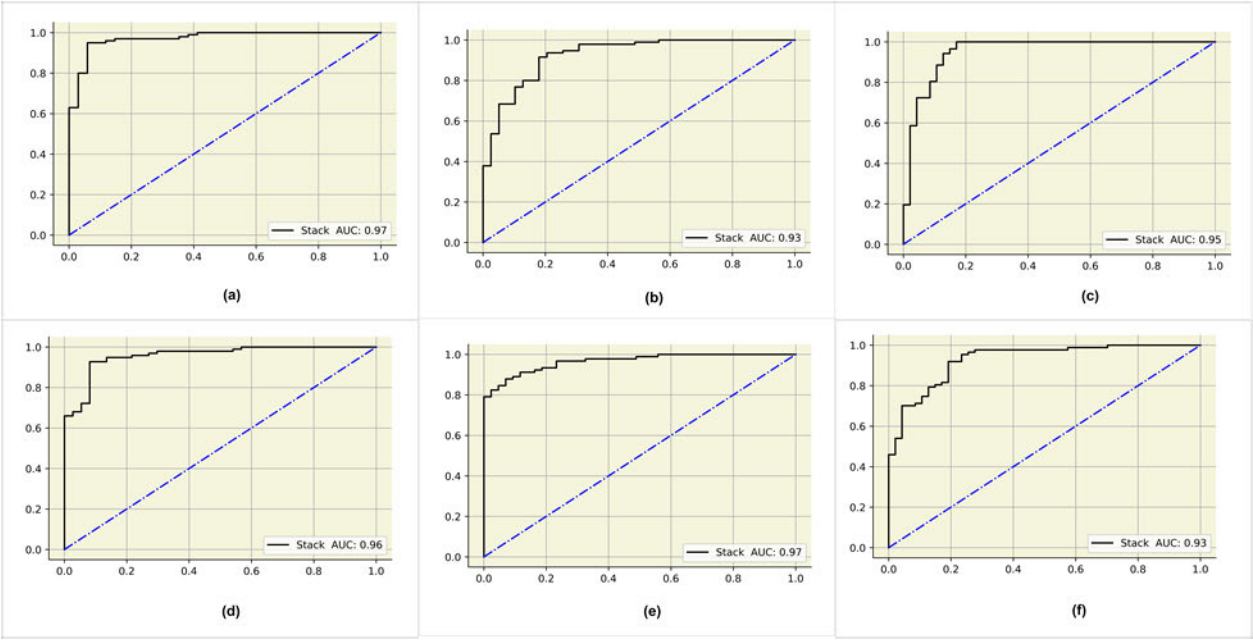

FIGURE 8. ROC curves for the optimal six feature selection algorithms. (a) MI (b) PC (c) BA (d) FPA (e) PSO and (f) SSO.

"Abzu" and is used sparingly in machine learning and data science. The models are demystified using QGraphs as shown in Fig. 15. These graphs contain activation functions, nodes and edges. Each attribute is represented using a node and the edges connect the nodes. The activation function is used to activate/deactivate the nodes. The figure shows that Albumin, TWBC, Sodium, Eosinophil and protein are the most essential markers considered by the Qgraphs. Activation functions such as inverse, tanh, addition, multiplication and Gaussian have been used for interpretation.

Anchor is an explainer that uses conditions to demystify the model outputs [56]. The explanations are further validated using two parameters: Coverage and precision. Precision measures the importance of the explanation, whereas coverage measures the number of samples which use the exact anchor explanation. The anchor conditions are presented in Table 9. Important markers according to Anchor, are Albumin, Protein, TWBC, T. Bilirubin and Eosinophil. The Albumin and Protein count increase in the COVID-19 cohort. The TWBC, T. Bilirubin, and Eosinophil levels slightly reduced in COVID-19 cases.

Five XAI methodologies were used for interpretation in this research. According to them, the crucial clinical attributes are Albumin, Protein, HbA1c, TWBC, Lymphocyte, ALP, Sodium and Eosinophils. Combining these markers and machine learning could be useful to for COVID-19 diagnosis.

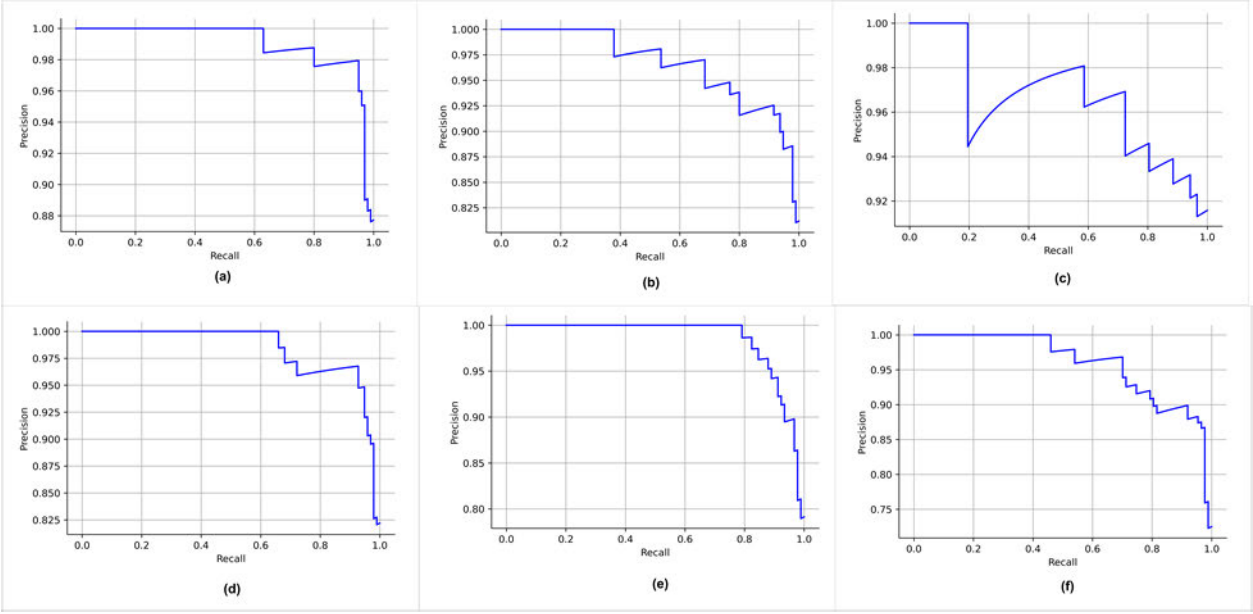

FIGURE 9. Precision-Recall curves for the optimal six feature selection algorithms. (a) MI (b) PC (c) BA (d) FPA (e) PSO and (f) SSO.

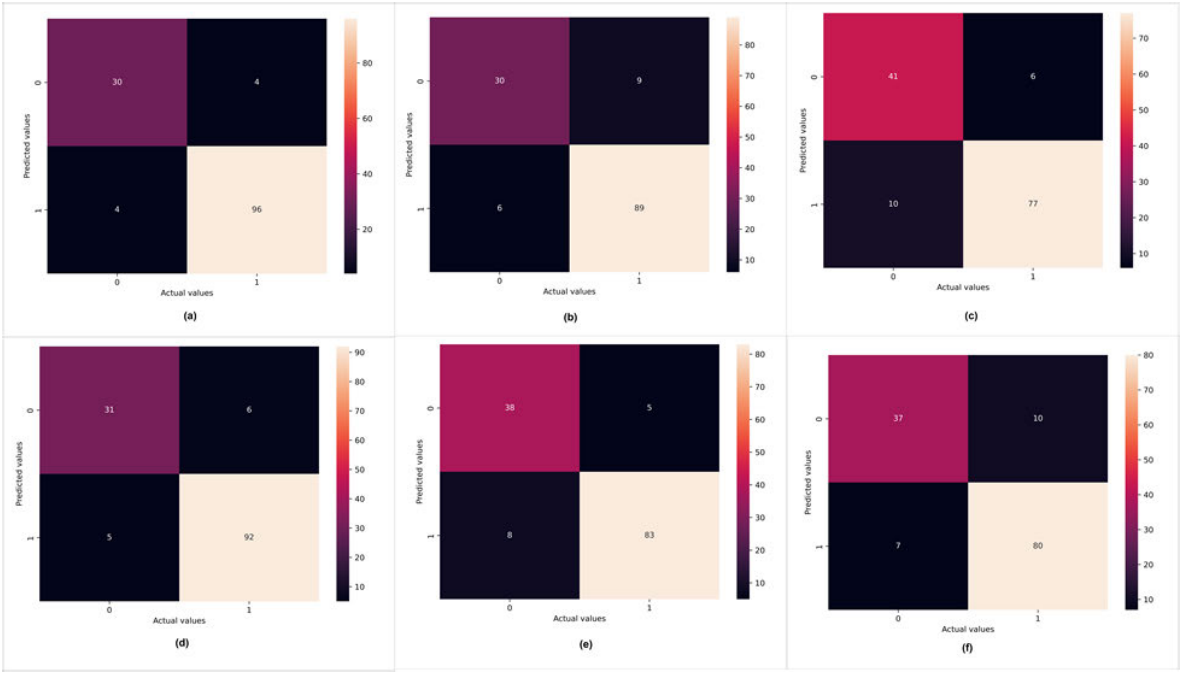

FIGURE 10. Confusion matrices for the optimal six feature selection algorithms. (a) MI (b) PC (c) BA (d) FPA (e) PSO and (f) SSO.

TABLE 8. Metrics obtained for deep learning models.

Classifier	Accuracy	Precision		Recall		F1-score		Hamming loss	Log loss	Jaccard score	MCC
		COVID-19	Non-COVID	COVID-19	Non-COVID-19	COVID-19	Non-COVID-19				
DNN	74%	91%	51%	71%	81%	62%	81%	0.26	9.02	0.66	0.46
1D-CNN	83%	89%	68%	88%	71%	88%	69%	0.16	5.82	0.79	0.57

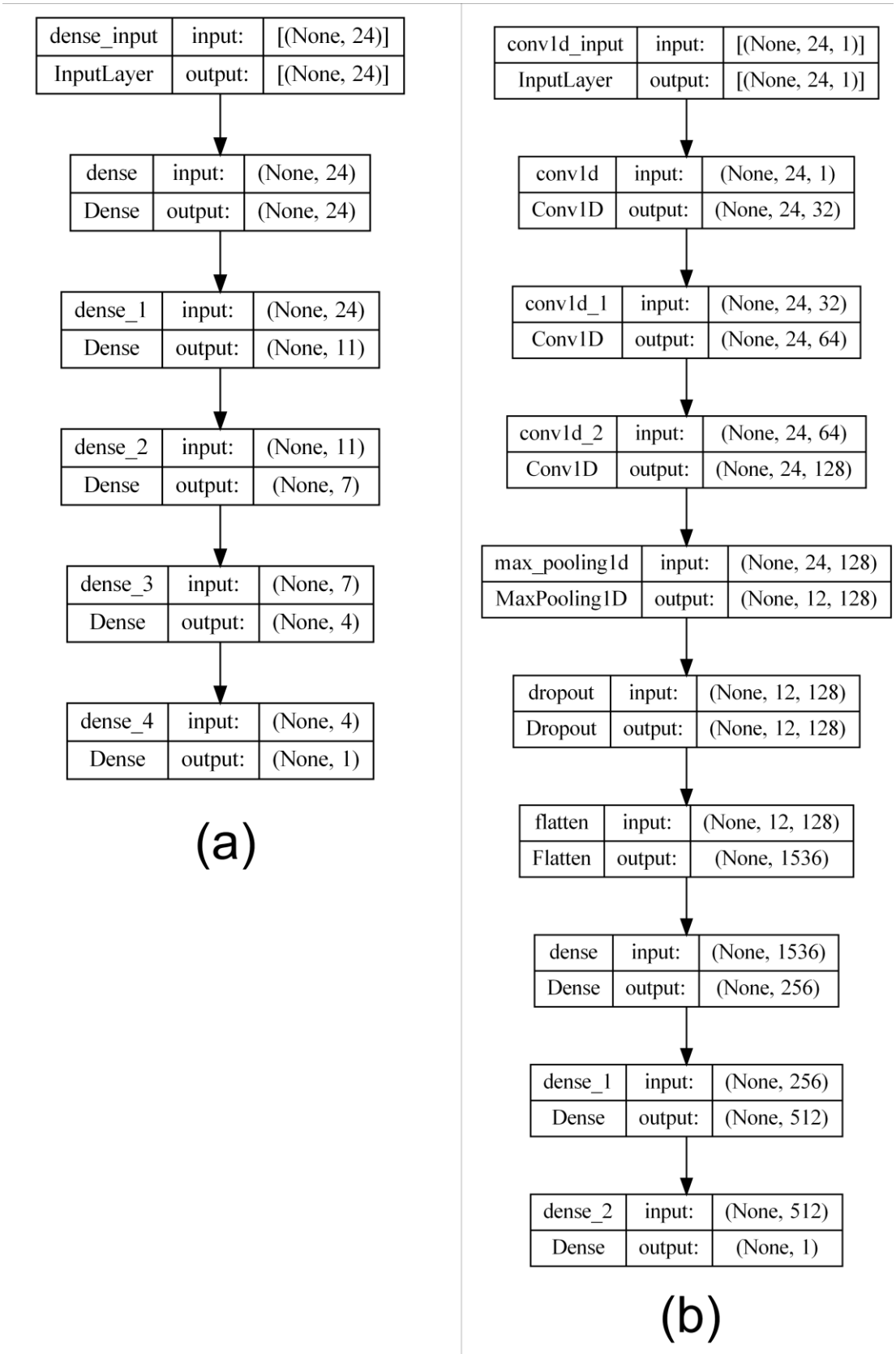

FIGURE 11. Architecture of the deep neural networks. (a) DNN (b) 1D-CNN.

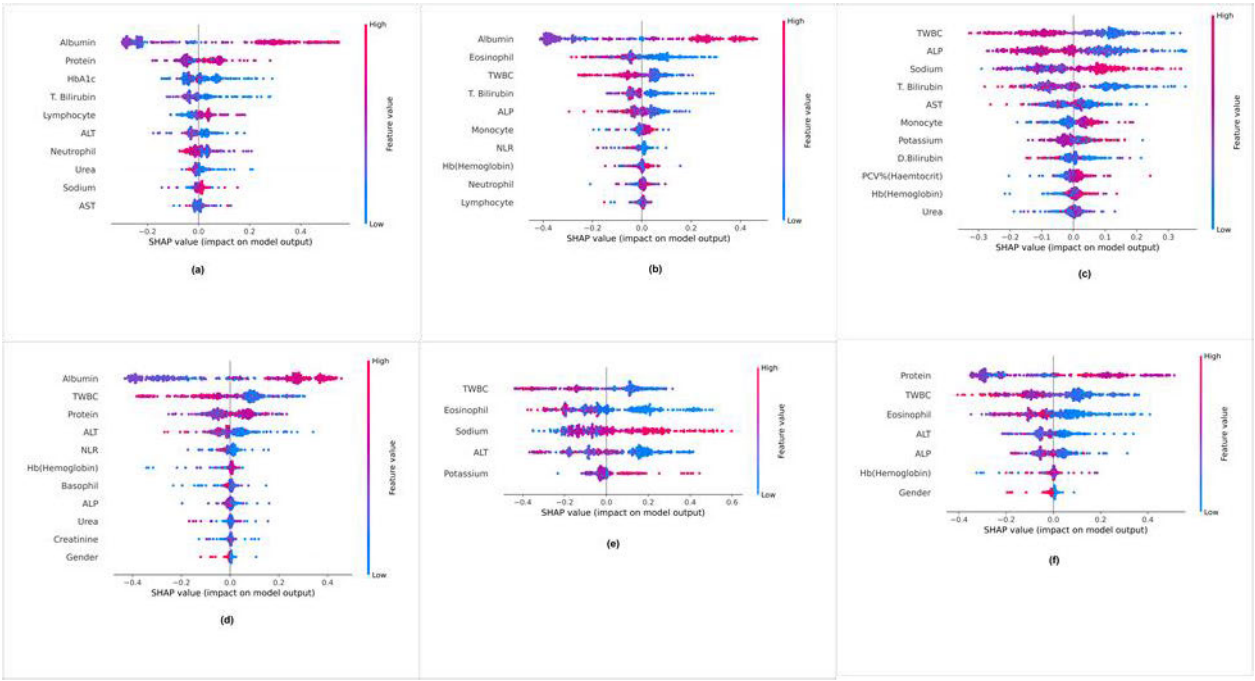

FIGURE 12. SHAP beeswarm plots. (a) MI (b) PC (c) BA (d) FPA (e) PSO and (f) SSO.

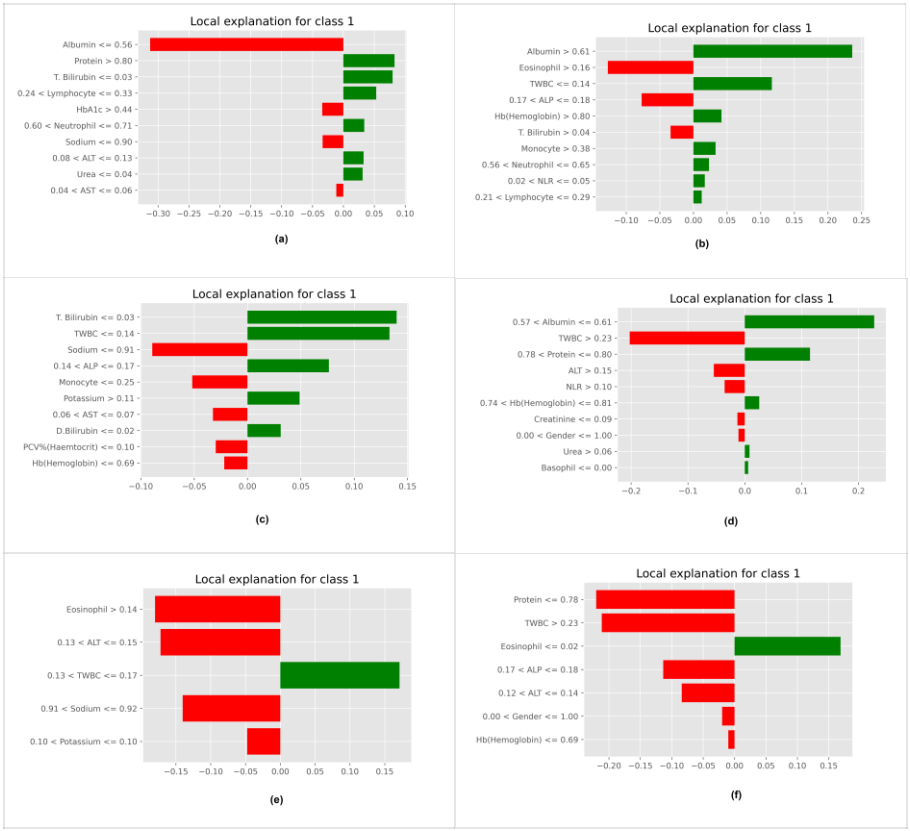

FIGURE 13. LIME plots for the optimal six feature selection algorithms. (a) MI (b) PC (c) BA (d) FPA (e) PSO and (f) SSO.

y=0.0 (probability **1.000**) top features

Contribution?	Feature	Value
+0.500	<BIAS>	1.000
+0.394	Albumin	0.557
+0.098	Protein	0.778
+0.007	HbA1c	0.386

(a)

y=1.0 (probability **1.000**) top features

Contribution?	Feature	Value
+0.500	<BIAS>	1.000
+0.451	Albumin	0.614
+0.041	Lymphocyte	0.418
+0.009	TWBC	0.131

(b)

y=1.0 (probability **1.000**) top features

Contribution?	Feature	Value
+0.500	Hb(Hemoglobin)	0.874
+0.500	<BIAS>	1.000
+0.372	Potassium	0.095
+0.239	ALP	0.137
-0.043	Sodium	0.926
-0.121	D.Bilirubin	0.010
-0.164	T. Bilirubin	0.048
-0.283	TWBC	0.222

(c)

y=1.0 (probability **1.000**) top features

Contribution?	Feature	Value
+0.500	<BIAS>	1.000
+0.333	Urea	0.026
+0.332	Albumin	0.614
+0.107	TWBC	0.120
-0.273	ALT	0.677

(d)

y=0.0 (probability **1.000**) top features

Contribution?	Feature	Value
+1.208	TWBC	0.214
+0.500	<BIAS>	1.000
-0.287	Sodium	0.932
-0.421	Eosinophil	0.007

(e)

y=1.0 (probability **1.000**) top features

Contribution?	Feature	Value
+0.500	<BIAS>	1.000
+0.491	Protein	0.800
+0.009	TWBC	0.140

(f)

FIGURE 14. Eli5 explanations for the optimal six feature selection algorithms. (a) MI (b) PC (c) BA (d) FPA (e) PSO and (f) SSO.

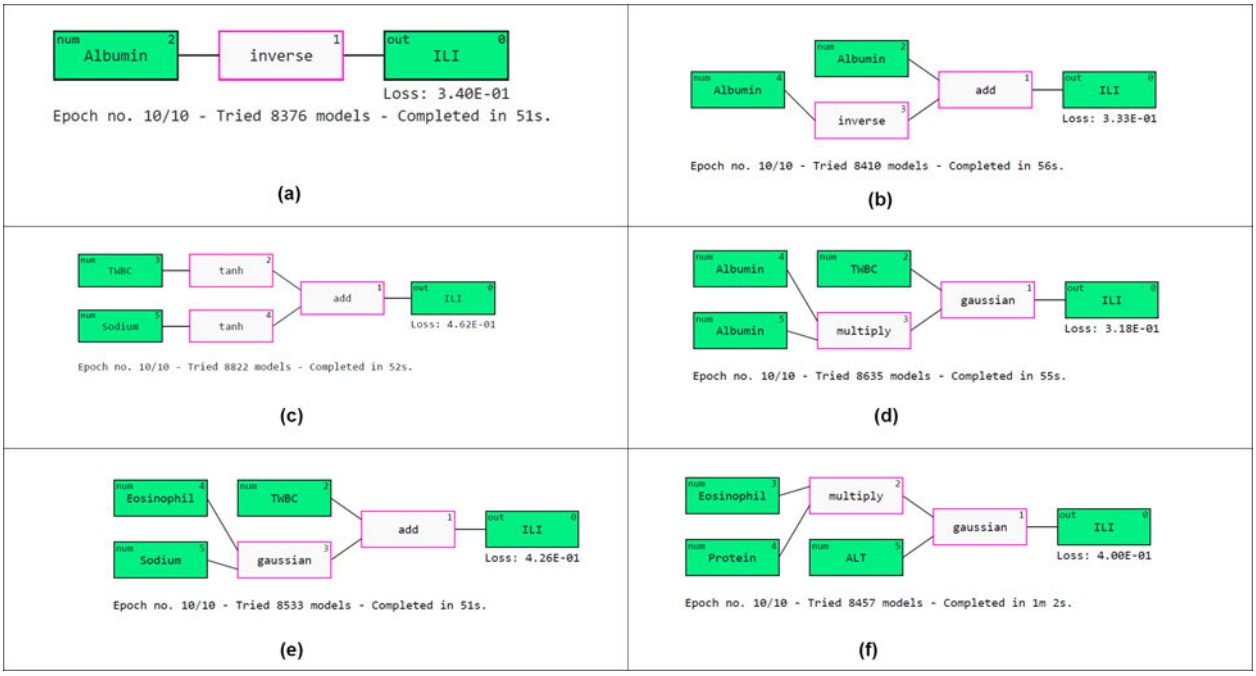

FIGURE 15. QGraphs for the optimal six feature selection algorithms. (a) MI (b) PC (c) BA (d) FPA (e) PSO and (f) SSO.

IV. DISCUSSION

In this retrospective research, ML algorithms were utilized to predict COVID-19 in patients. The patient data was collected from two medical facilities in India after taking appropriate ethical permissions to conduct this research. Twenty-four markers were analyzed statistically in the beginning. A few data preprocessing steps were conducted later on the data. ALP, ALT, Hemoglobin and TWBC were the highest chosen markers by the feature selection methods. Various machine learning algorithms were tested later, including stacking and voting classifiers. Two deep neural network models were also used for COVID-19 diagnosis. Several classification and loss metrics were used for performance evaluation. The predictions made by the algorithms were deciphered using five explainers. According to them, the crucial markers are Albumin, Protein, HbA1c, TWBC, Lymphocyte, ALP, Sodium and Eosinophils.

The top six feature selection algorithms were considered for in-depth analysis. The mutual algorithm obtained the optimal results among all the techniques. Two boosting

TABLE 9. Model explanations using anchor.

Feature selection method	Diagnosis	Condition (Anchor explanation)	Precision of the anchor condition	Coverage of the anchor condition
Mutual information	COVID-19 positive diagnosis	Albumin > 0.56 and 0.78 < Protein <=0.80	0.92	0.37
	COVID-19 negative diagnosis	Albumin < 0.56 and Protein > 0.78	0.92	0.38
Pearson's correlation	COVID-19 positive diagnosis	Albumin > 0.57 and TWBC <=0.14	0.94	0.16
	COVID-19 negative diagnosis	Albumin <=0.56 and Eosinophil > 0.16	0.93	0.13
Bat algorithm	COVID-19 positive diagnosis	TWBC > 0.17 and T. Bilirubin > 0.03	0.70	0.35
	COVID-19 negative diagnosis	Sodium > 0.94 and T. Bilirubin <=0 0.03	0.83	0.10
Flower pollination algorithm	COVID-19 positive diagnosis	Albumin > 0.57 and TWBC <= 0.18	0.90	0.31
	COVID-19 negative diagnosis	Albumin <=0.56 and TWBC > 0.18	0.95	0.42
Particle swarm optimization	COVID-19 positive diagnosis	Eosinophil <= 0.01 and Sodium > 0.94	0.97	0.03
	COVID-19 negative diagnosis	TWBC > 0.23 and Sodium <=0.91	0.86	0.06
Salp swarm optimization	COVID-19 positive diagnosis	TWBC <=0.14 and Protein > 0.78	0.88	0.17
	COVID-19 negative diagnosis	TWBC > 0.18 and Protein <=0.78	0.90	0.21

models were able to attain an accuracy of 95%. Among the deep learning neural networks, 1D-CNN performed well with an accuracy of 83%. After using five heterogeneous techniques, we identified the most important markers for COVID-19 diagnosis.

Albumin levels were raised in mild COVID-19 cases. Lower levels of albumin can indicate damage to the kidneys/liver [57]. Higher Protein levels were also observed in coronavirus patients. Lower Protein levels can lead to Hypoproteinemia [58]. HbA1c levels were also lower in mild coronavirus patients. Elevated HbA1c levels were found in moderate and severe COVID-19 patients, according to many studies [59], [60]. Neutrophil levels are known to elevate with a severe infection [61]. The neutrophil count was lesser in the mild COVID-19 class. Many researchers have also observed decreased lymphocyte count in COVID-19 patients with dire prognoses [62]. Our study agrees, and the lymphocyte count was elevated in the COVID-19 group. Neutrophil to lymphocyte levels were lesser in the mild COVID-19 cohort and agree with other similar published studies [63], [64]. Liver biomarkers such as D. Bilirubin, T. Bilirubin, AST, ALT and ALP were elevated in the non-COVID-19 group [65].

Many articles have been published that use clinical markers to diagnose COVID-19. Baik et al. [20] used boosting models to conduct differential diagnosis among COVID-19 and other patients. Model explainability was achieved using SHAP. According to the study, the most essential features were D-Dimer, Glucose, Eosinophil, Basophil and AST. Haematological markers were used to detect COVID-19 influenza among children [19]. The machine learning models obtained a maximum accuracy of 98.4%. Using standardized blood tests, a deep convolutional classifier was used for COVID-19 diagnosis [14]. An accuracy of 98.5% was attained during testing. The critical markers were LDH, WBC and AST. Mohammadi et al. [16] evaluated the effectiveness of multilayer perceptron in diagnosing COVID-19 using laboratory markers. The algorithm obtained an accuracy of 99.16%. CRP, lymphocytes, magnesium and vitamin D were considered more important. The number of COVID- 19 studies which conducted differential diagnosis using machine learning and clinical markers was very few.

However, a few limitations exist in this study. While biomarkers are helpful, they cannot provide a complete prediction and may have their limitations and variations [66]. Other diagnostic procedures such as X-rays, CT-Scans and antigen-based testing could be suitably combined. The data considered in this study was obtained from two medical facilities in India. COVID-19 datasets from different Countries could be suitably combined in the future. The size of the dataset could also be increased for better analysis. In this study, a few algorithms were not explored. Therefore, unsupervised and reinforcement learning algorithms could also be tested. Cloud frameworks were not used for storage in this study. This can lead to memory/space issues when the data is significant. Hence, cloud-based architectures could be deployed in the future. Deep learning models perform well when the data is enormous. Hence, in the future, the sample size could be more. A user interface could be developed in the future. This module would make it easier for medical professionals to test the classifiers. The models were not deployed in real-time in Hospitals. Clinical validation could be performed, and the model predictions could be compared

with the treatment outcomes. Emphasis should also be given to data security and privacy. Sensitive details such as patient information should not be leaked. Other health monitors such as oxygen saturation, blood pressure pulse rate and heart rate could also be considered.

V. CONCLUSION

COVID-19 vaccines were critical in curbing the severe symptoms caused by the virus. However, this contagious virus continues to exist along with other similar illnesses. Hence, supervised artificial intelligence techniques are utilized in this retrospective study to predict mild COVID-19 from other influenzas. Multiple machine learning and deep neural networks were trained and used for prediction. Among the models, stacking and voting classifiers proved to be the best. Five XAI techniques were also used to interpret the best-performing classifiers. According to them, the best markers required for differential diagnosis were Albumin, Protein, HbA1c, TWBC, Lymphocytes, ALP, Sodium and Eosinophils. The albumin, protein, lymphocytes and sodium were higher in mild COVID-19 cohort. TWBC, ALP and Eosinophil were lower in the mild COVID-19 cohort. This model architecture could be deployed as a decision support system to identify mild SARS-CoV-2 patients. The models can also ease the load on healthcare personnel and medical infrastructure.

In the future, datasets from various hospitals could be combined to make the results more reliable. A user-interface could be created for real time prediction of COVID-19 patients. The data and models could be deployed on a cloud platform. Other demographic, immunological, clinical and epidemiological markers could also be combined. Other diagnostic methods such as Chest X-rays and CT Scans could also be used synchronously to make the predictions better.

A. ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Ethical clearance has been obtained to collect patient data from Manipal Academy of Higher Education ethics committee with id IEC: 613/2021. The need for informed consent was waived by the ethics committee/Institutional Review Board of Manipal Academy of Higher Education, because of the retrospective nature of the study. All methods were carried out in accordance with relevant guidelines and regulations.

REFERENCES

[1] R. T. Gandhi, J. B. Lynch, and C. del Rio, "Mild or moderate COVID-19," *New England J. Med.*, vol. 383, no. 18, pp. 1757–1766, Oct. 2020, doi: 10.1056/nejmcp2009249.

[2] G. Pascarella, A. Strumia, C. Piliego, F. Bruno, R. Del Buono, F. Costa, S. Scarlata, and F. E. Agro, "COVID-19 diagnosis and management: A comprehensive review," *J. Internal Med.*, vol. 288, no. 2, pp. 192–206, Aug. 2020, doi: 10.1111/joim.13091.

[3] Y. Hirotsu, M. Maejima, M. Shibusawa, K. Amemiya, Y. Nagakubo, K. Hosaka, H. Sueki, H. Mochizuki, T. Tsutsui, Y. Kakizaki, Y. Miyashita, and M. Omata, "Analysis of COVID-19 and non-COVID-19 viruses, including influenza viruses, to determine the influence of intensive preventive measures in Japan," *J. Clin. Virol.*, vol. 129, Aug. 2020, Art. no. 104543, doi: 10.1016/j.jcv.2020.104543.

[4] A. Scohy, A. Anantharajah, M. Bodéus, B. Kabamba-Mukadi, A. Verroken, and H. Rodriguez-Villalobos, "Low performance of rapid antigen detection test as frontline testing for COVID-19 diagnosis," *J. Clin. Virol.*, vol. 129, Aug. 2020, Art. no. 104455, doi: 10.1016/j.jcv.2020.104455.

[5] M. Nagura-Ikeda, K. Imai, S. Tabata, K. Miyoshi, N. Murahara, T. Mizuno, M. Horiuchi, K. Kato, Y. Imoto, M. Iwata, S. Mimura, T. Ito, K. Tamura, and Y. Kato, "Clinical evaluation of self-collected saliva by quantitative reverse transcription-PCR (RT-qPCR), direct RT-qPCR, reverse transcription-loop-mediated isothermal amplification, and a rapid antigen test to diagnose COVID-19," *J. Clin. Microbiol.*, vol. 58, no. 9, Aug. 2020, doi: 10.1128/jcm.01438-20.

[6] G. Liu and J. F. Rusling, "COVID-19 antibody tests and their limitations," *ACS Sensors*, vol. 6, no. 3, pp. 593–612, Feb. 2021, doi: 10.1021/acssensors.0c02621.

[7] O. Filchakova, D. Dossym, A. Ilyas, T. Kuanysheva, A. Abdizhamil, and R. Bukasov, "Review of COVID-19 testing and diagnostic methods," *Talanta*, vol. 244, Jul. 2022, Art. no. 123409, doi: 10.1016/j.talanta.2022.123409.

[8] L. Rubinger, A. Gazendam, S. Ekhtiari, and M. Bhandari, "Machine learning and artificial intelligence in research and healthcare," *Injury*, vol. 54, pp. S69–S73, May 2023, doi: 10.1016/j.injury.2022.01.046.

[9] K. Santosh and L. Gaur, "Introduction to AI in public health," in *Artificial Intelligence and Machine Learning in Public Healthcare*. Singapore: Springer, 2021, pp. 1–10, doi: 10.1007/978-981-16-6768-8_1.

[10] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," 2019, *arXiv:1910.10045*.

[11] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018, doi: 10.1109/ACCESS.2018.2870052.

[12] B. Gallo Marin, G. Aghagoli, K. Lavine, L. Yang, E. J. Siff, S. S. Chiang, T. P. Salazar-Mather, L. Dumenco, M. C. Savaria, S. N. Aung, T. Flanigan, and I. C. Michelow, "Predictors of COVID-19 severity: A literature review," *Rev. Med. Virol.*, vol. 31, no. 1, pp. 1–10, Jul. 2020, doi: 10.1002/rmv.2146.

[13] H. Gabr, S. Bastawy, A. A. Abdel_Aal, N. M. Khalil, and M. Fateen, "Changes in peripheral blood cellular morphology as diagnostic markers for COVID-19 infection," *Int. J. Lab. Hematol.*, vol. 44, no. 3, pp. 454–460, Jan. 2022, doi: 10.1111/ijlh.13799.

[14] D. A. Altantawy and S. S. Kishk, "Equilibrium-based COVID-19 diagnosis from routine blood tests: A sparse deep convolutional model," *Expert Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 118935, doi: 10.1016/j.eswa.2022.118935.

[15] A. A. D. Souza, D. C. D. Almeida, T. S. Barcelos, R. C. Bortoletto, R. Munoz, H. Waldman, M. A. Goes, and L. A. Silva, "Simple hemogram to support the decision-making of COVID-19 diagnosis using clusters analysis with self-organizing maps neural network," *Soft Comput.*, vol. 27, no. 6, pp. 3295–3306, May 2021, doi: 10.1007/s00500-021-05810-5.

[16] F. Mohammadi, L. Dehbozorgi, H. R. Akbari-Hasanjani, Z. J. Abbasalian, R. Akbari-Hasanjani, R. Sabbaghi-Nadooshan, and H. M. Tabriz, "Evaluation of effective features in the diagnosis of COVID-19 infection from routine blood tests with multilayer perceptron neural network: A cross-sectional study," *Health Sci. Rep.*, vol. 6, no. 1, Jan. 2023, doi: 10.1002/hsr2.1048.

[17] S. Solayman, S. A. Aumi, C. S. Mery, M. Mubassir, and R. Khan, "Automatic COVID-19 prediction using explainable machine learning techniques," *Int. J. Cognit. Comput. Eng.*, vol. 4, pp. 36–46, Jun. 2023, doi: 10.1016/j.ijcce.2023.01.003.

[18] T. Roland, C. Böck, T. Tschoellitsch, A. Maletzky, S. Hochreiter, J. Meier, and G. Klambauer, "Domain shifts in machine learning based COVID-19 diagnosis from blood tests," *J. Med. Syst.*, vol. 46, no. 5, Mar. 2022, doi: 10.1007/s10916-022-01807-1.

[19] D. Dobrijević, J. Antić, G. Rakić, J. Katanić, L. Andrijević, and K. Pastor, "Clinical hematochemical parameters in differential diagnosis between pediatric SARS-CoV-2 and influenza virus infection: An automated machine learning approach," *Children*, vol. 10, no. 5, p. 761, Apr. 2023, doi: 10.3390/children10050761.

[20] S. M. Baik, K. S. Hong, and D. J. Park, "Application and utility of boosting machine learning model based on laboratory test in the differential diagnosis of non-COVID-19 pneumonia and COVID-19," *Clin. Biochem.*, vol. 118, Aug. 2023, Art. no. 110584, doi: 10.1016/j.clinbiochem.2023.05.003.

[21] F. Styrzynski, D. Zhakparov, M. Schmid, D. Roqueiro, Z. Lukasik, J. Solek, J. Nowicki, M. Dobrogowski, J. Makowska, M. Sokolowska, and K. Baerenfaller, "Machine learning successfully detects patients with COVID-19 prior to PCR results and predicts their survival based on standard laboratory parameters in an observational study," *Infectious Diseases Therapy*, vol. 12, no. 1, pp. 111–129, Nov. 2022, doi: 10.1007/s40121-022-00707-8.

[22] J. C. Gomes, V. A. de Freitas Barbosa, M. A. de Santana, C. L. de Lima, R. B. Calado, C. R. B. Júnior, J. E. de Almeida Albuquerque, R. G. de Souza, R. J. E. de Araújo, G. M. M. Moreno, L. A. L. Soares, L. A. R. M. Júnior, R. E. de Souza, and W. P. dos Santos, "Rapid protocols to support COVID-19 clinical diagnosis based on hematological parameters," *Res. Biomed. Eng.*, vol. 39, no. 3, pp. 509–539, Jun. 2023, doi: 10.1007/s42600-023-00286-8.

[23] M. Şahin and E. Aybek, "Jamovi: An easy to use statistical software for the social scientists," *Int. J. Assessment Tools Educ.*, vol. 6, no. 4, pp. 670–692, Jan. 2020, doi: 10.21449/ijate.661803.

[24] G. Gunč ar, M. Kukar, M. Notar, M. Brvar, P. Černelč, M. Notar, and M. Notar, "An application of machine learning to haematological diagnosis," *Sci. Rep.*, vol. 8, no. 1, Jan. 2018, doi: 10.1038/s41598-017-18564-8.

[25] N. Kosaraju, S. R. Sankepally, and K. M. Rao, "Categorical data: Need, encoding, selection of encoding method and its emergence in machine learning models—A practical review study on heart disease prediction dataset using Pearson correlation," in *Proc. Int. Conf. Data Sci. Appl.*, in Lecture notes in networks and systems, Jan. 2023, pp. 369–382, doi: 10.1007/978-981-19-6631-6_26.

[26] P. Rodríguez, M. A. Bautista, J. Gonzàlez, and S. Escalera, "Beyond one-hot encoding: Lower dimensional target embedding," *Image Vis. Comput.*, vol. 75, pp. 21–31, Jul. 2018, doi: 10.1016/j.imavis.2018.04.004.

[27] Y. S. Solanki, P. Chakrabarti, M. Jasinski, Z. Leonowicz, V. Bolshev, A. Vinogradov, E. Jasinska, R. Gono, and M. Nami, "A hybrid supervised machine learning classifier system for breast cancer prognosis using feature selection and data imbalance handling approaches," *Electronics*, vol. 10, no. 6, p. 699, Mar. 2021, doi: 10.3390/electronics10060699.

[28] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning," in *Advances in Intelligent Computing* (Lecture Notes in Computer Science), vol. 3644. Berlin, Germany: Springer, 2005, pp. 878–887, doi: 10.1007/11538059_91.

[29] D. U. Ozsahin, M. T. Mustapha, A. S. Mubarak, Z. S. Ameen, and B. Uzun, "Impact of feature scaling on machine learning models for the diagnosis of diabetes," in *Proc. Int. Conf. Artif. Intell. Everything (AIE)*, Aug. 2022, pp. 87–94, doi: 10.1109/AIE57029.2022.00024.

[30] S. Kaur, Y. Kumar, A. Koul, and S. K. Kamboj, "A systematic review on metaheuristic optimization techniques for feature selections in disease diagnosis: Open issues and challenges," *Arch. Comput. Methods Eng.*, vol. 30, no. 3, pp. 1863–1895, Nov. 2022, doi: 10.1007/s11831-022-09853-1.

[31] A. M. Vommi and T. K. Battula, "A hybrid filter-wrapper feature selection using fuzzy KNN based on Bonferroni mean for medical datasets classification: A COVID-19 case study," *Expert Syst. Appl.*, vol. 218, May 2023, Art. no. 119612, doi: 10.1016/j.eswa.2023.119612.

[32] Z. Li, Y. Yang, L. Li, and D. Wang, "A weighted Pearson correlation coefficient based multi-fault comprehensive diagnosis for battery circuits," *J. Energy Storage*, vol. 60, Apr. 2023, Art. no. 106584, doi: 10.1016/j.est.2022.106584.

[33] X. Lei, Y. Xia, A. Wang, X. Jian, H. Zhong, and L. Sun, "Mutual information based anomaly detection of monitoring data with attention mechanism and residual learning," *Mech. Syst. Signal Process.*, vol. 182, Jan. 2023, Art. no. 109607, doi: 10.1016/j.ymssp.2022.109607.

[34] A. Yaqoob, R. M. Aziz, N. K. Verma, P. Lalwani, A. Makrariya, and P. Kumar, "A review on nature-inspired algorithms for cancer disease prediction and classification," *Mathematics*, vol. 11, no. 5, p. 1081, Feb. 2023, doi: 10.3390/math11051081.

[35] Y. Sun, S. Ding, Z. Zhang, and W. Jia, "An improved grid search algorithm to optimize SVR for prediction," *Soft Comput.*, vol. 25, no. 7, pp. 5633–5644, Jan. 2021, doi: 10.1007/s00500-020-05560-w.

[36] B. Pavlyshenko. (Aug. 1, 2018). *Using Stacking Approaches for Machine Learning Models*. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8478522

[37] D. Burka, C. Puppe, L. Szepesváry, and A. Tasnádi, "Voting: A machine learning approach," *Eur. J. Oper. Res.*, vol. 299, no. 3, pp. 1003–1017, Jun. 2022, doi: 10.1016/j.ejor.2021.10.005.

[38] G. Montavon, W. Samek, and K.-R. Müller, "Methods for interpreting and understanding deep neural networks," *Digit. Signal Process.*, vol. 73, pp. 1–15, Feb. 2018, doi: 10.1016/j.dsp.2017.10.011.

[39] L. Eren, T. Ince, and S. Kiranyaz, "A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier," *J. Signal Process. Syst.*, vol. 91, no. 2, pp. 179–189, May 2018, doi: 10.1007/s11265-018-1378-3.

[40] H. Prabowo, A. A. Hidayat, T. W. Cenggoro, R. Rahutomo, K. Purwandari, and B. Pardamean, "Aggregating time series and tabular data in deep learning model for university students' GPA prediction," *IEEE Access*, vol. 9, pp. 87370–87377, 2021, doi: 10.1109/ACCESS.2021.3088152.

[41] H. Touvron, P. Bojanowski, M. Caron, M. Cord, A. El-Nouby, E. Grave, G. Izacard, A. Joulin, G. Synnaeve, J. Verbeek, and H. Jégou, "ResMLP: Feedforward networks for image classification with data-efficient training," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 5314–5321, Apr. 2023, doi: 10.1109/TPAMI.2022.3206148.

[42] S. Eskandari and M. Seifaddini, "Online and offline streaming feature selection methods with bat algorithm for redundancy analysis," *Pattern Recognit.*, vol. 133, Jan. 2023, Art. no. 109007, doi: 10.1016/j.patcog.2022.109007.

[43] B. Sahu, P. K. Das, and R. Kumar, "A modified cuckoo search algorithm implemented with SCA and PSO for multi-robot cooperation and path planning," *Cogn. Syst. Res.*, vol. 79, pp. 24–42, Jun. 2023, doi: 10.1016/j.cogsys.2023.01.005.

[44] B. Mohammadi, "Modeling various drought time scales via a merged artificial neural network with a firefly algorithm," *Hydrology*, vol. 10, no. 3, p. 58, Feb. 2023, doi: 10.3390/hydrology10030058.

[45] H. Zhang, J. Gao, L. Kang, Y. Zhang, L. Wang, and K. Wang, "State of health estimation of lithium-ion batteries based on modified flower pollination algorithm-temporal convolutional network," *Energy*, vol. 283, Nov. 2023, Art. no. 128742, doi: 10.1016/j.energy.2023.128742.

[46] A. Sohail, "Genetic algorithms in the fields of artificial intelligence and data sciences," *Ann. Data Sci.*, vol. 10, no. 4, pp. 1007–1018, Aug. 2021, doi: 10.1007/s40745-021-00354-9.

[47] S. N. Makhadmeh, M. A. Al-Betar, I. A. Doush, M. A. Awadallah, S. Kassaymeh, S. Mirjalili, and R. A. Zitar, "Recent advances in grey wolf optimizer, its versions and applications: Review," *IEEE Access*, vol. 12, pp. 22991–23028, 2024, doi: 10.1109/ACCESS.2023.3304889.

[48] H. M. Alabool, D. Alarabiat, L. Abualigah, and A. A. Heidari, "Harris hawks optimization: A comprehensive review of recent variants and applications," *Neural Comput. Appl.*, vol. 33, no. 15, pp. 8939–8980, Feb. 2021, doi: 10.1007/s00521-021-05720-5.

[49] N. K. Jain, U. Nangia, and J. Jain, "A review of particle swarm optimization," *J. Inst. Eng. India, B*, vol. 99, no. 4, pp. 407–411, Mar. 2018, doi: 10.1007/s40031-018-0323-y.

[50] M. Castelli, L. Manzoni, L. Mariot, M. S. Nobile, and A. Tangherloni, "Salp swarm optimization: A critical review," *Expert Syst. Appl.*, vol. 189, Mar. 2022, Art. no. 116029, doi: 10.1016/j.eswa.2021.116029.

[51] N. Rana, M. S. A. Latiff, S. M. Abdulhamid, and H. Chiroma, "Whale optimization algorithm: A systematic review of contemporary applications, modifications and developments," *Neural Comput. Appl.*, vol. 32, no. 20, pp. 16245–16277, Mar. 2020, doi: 10.1007/s00521-020-04849-z.

[52] P. Gupta, S. Maji, and R. Mehra, "Predictive modeling of stress in the healthcare industry during COVID-19: A novel approach using XGBoost, SHAP values, and tree explainer," *Int. J. Decis. Support Syst. Technol.*, vol. 15, no. 1, pp. 1–20, Jan. 2023, doi: 10.4018/ijdsst.315758.

[53] H. W. Loh, C. P. Ooi, S. Seoni, P. D. Barua, F. Molinari, and U. R. Acharya, "Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022)," *Comput. Methods Programs Biomed.*, vol. 226, Nov. 2022, Art. no. 107161, doi: 10.1016/j.cmpb.2022.107161.

[54] M. S. Islam, I. Hussain, M. M. Rahman, S. J. Park, and M. A. Hossain, "Explainable artificial intelligence model for stroke prediction using EEG signal," *Sensors*, vol. 22, no. 24, p. 9859, Dec. 2022, doi: 10.3390/s22249859.

[55] C. Hu, C. Gao, T. Li, C. Liu, and Z. Peng, "Explainable artificial intelligence model for mortality risk prediction in the intensive care unit: A derivation and validation study," *Postgraduate Med. J.*, vol. 100, no. 1182, pp. 219–227, Jan. 2024, doi: 10.1093/postmj/qgad144.

[56] S. S. Band, A. Yarahmadi, C.-C. Hsu, M. Biyari, M. Sookhak, R. Ameri, I. Dehzangi, A. T. Chronopoulos, and H.-W. Liang, "Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods," *Informat. Med. Unlocked*, vol. 40, Jan. 2023, Art. no. 101286, doi: 10.1016/j.imu.2023.101286.

[57] R. de la Rica, M. Borges, M. Aranda, A. del Castillo, A. Socias, A. Payeras, G. Rialp, L. Socias, L. Masmiquel, and M. Gonzalez-Freire, "Low albumin levels are associated with poorer outcomes in a case series of COVID-19 patients in spain: A retrospective cohort study," *Microorganisms*, vol. 8, no. 8, p. 1106, Jul. 2020, doi: 10.3390/microorganisms8081106.

[58] S. Hojyo, M. Uchida, K. Tanaka, R. Hasebe, Y. Tanaka, M. Murakami, and T. Hirano, "How COVID-19 induces cytokine storm with high mortality," *Inflammation Regeneration*, vol. 40, no. 1, pp. 1–7, Oct. 2020, doi: 10.1186/s41232-020-00146-3.

[59] Z. Zhu, Y. Mao, and G. Chen, "Predictive value of HbA1c for in-hospital adverse prognosis in COVID-19: A systematic review and meta-analysis," *Primary Care Diabetes*, vol. 15, no. 6, pp. 910–917, Dec. 2021, doi: 10.1016/j.pcd.2021.07.013.

[60] F. Prattichizzo, P. de Candia, A. Nicolucci, and A. Ceriello, "Elevated HbA1c levels in pre-COVID-19 infection increases the risk of mortality: A systematic review and meta-analysis," *Diabetes/Metabolism Res. Rev.*, vol. 38, no. 1, May 2021, doi: 10.1002/dmrr.3476.

[61] G. D. Pimentel, M. C. M. Dela Vega, and A. Laviano, "High neutrophil to lymphocyte ratio as a prognostic marker in COVID-19 patients," *Clin. Nutrition ESPEN*, vol. 40, pp. 101–102, Dec. 2020, doi: 10.1016/j.clnesp.2020.08.004.

[62] J. R. Ulloque-Badaracco, W. Ivan Salas-Tello, A. Al-kassab-Córdova, E. A. Alarcón-Braga, V. A. Benites-Zapata, J. L. Maguiña, and A. V. Hernandez, "Prognostic value of neutrophil-to-lymphocyte ratio in COVID-19 patients: A systematic review and meta-analysis," *Int. J. Clin. Pract.*, vol. 75, no. 11, Jul. 2021, doi: 10.1111/ijcp.14596.

[63] Y. Liu, X. Du, J. Chen, Y. Jin, L. Peng, H. H. X. Wang, M. Luo, L. Chen, and Y. Zhao, "Neutrophil-to-lymphocyte ratio as an independent risk factor for mortality in hospitalized patients with COVID-19," *J. Infection*, vol. 81, no. 1, pp. e6–e12, Jul. 2020, doi: 10.1016/j.jinf.2020.04.002.

[64] X. Li, C. Liu, Z. Mao, M. Xiao, L. Wang, S. Qi, and F. Zhou, "Predictive values of neutrophil-to-lymphocyte ratio on disease severity and mortality in COVID-19 patients: A systematic review and meta-analysis," *Crit. Care*, vol. 24, no. 1, p. 647, Nov. 2020, doi: 10.1186/s13054-020-03374-8.

[65] Z. Fan, L. Chen, J. Li, X. Cheng, J. Yang, C. Tian, Y. Zhang, S. Huang, Z. Liu, and J. Cheng, "Clinical features of COVID-19-related liver functional abnormality," *Clin. Gastroenterol. Hepatol.*, vol. 18, no. 7, pp. 1561–1566, Jun. 2020, doi: 10.1016/j.cgh.2020.04.002.

[66] S. Der Sarkissian, S. Hessam, J. S. Kirby, M. A. Lowes, D. Mintoff, H. B. Naik, H. C. Ring, N. S. Chandran, and J. W. Frew, "Identification of biomarkers and critical evaluation of biomarker validation in hidradenitis suppurativa," *J. Amer. Med. Assoc. Dermatol.*, vol. 158, no. 3, p. 300, Mar. 2022, doi: 10.1001/jamadermatol.2021.4926.

KRISHNARAJ CHADAGA received the B.E. degree from the Department of Computer Science and Engineering, Shri Madwavadiraja Institute of Technology, and the M.Tech. degree from Manipal Institute of Technology, Manipal.

He is currently a Research Scholar with the Department of Computer Science and Engineering, Manipal Institute of Technology. Prior to joining Ph.D. degree, he worked in many companies, such as DELL EMC, Informatica, and Diya Systems. His research interests include medical informatics, disease classification, decision support management systems, and disease diagnosis/prognosis prediction. He is a peer-reviewer in many academic journals.

SRIKANTH PRABHU (Senior Member, IEEE) received the M.Sc., M.Tech., and Ph.D. degrees from IIT Kharagpur. He is currently a Professor with the Department of Computer Science and Engineering, Manipal Institute of Technology, MAHE, Manipal. He has more than 150 publications in national and international conferences and journals. His current research interests include pattern recognition, pattern classification, fuzzy logic, image processing, and parallel processing.

NIRANJANA SAMPATHILA (Senior Member, IEEE) received the B.E. degree in electronics and engineering and the M.Tech. degree in biomedical engineering from Mangalore University and the Ph.D. degree from Manipal University [currently known as Manipal Academy of Higher Education (MAHE)], Manipal, India. He is currently a Professor with Manipal Institute of Technology, MAHE. He has more than 27 years of experience in research and academics in biomedical engineering applications. His research interests include pattern recognition, medical informatics, miniaturized system design, and AI for healthcare. He is a fellow of IE, India.

RAJAGOPALA CHADAGA received the B.E. degree from the S. J. M. Institute of Technology, Chitradurga, and the M.Tech. degree from NIE, Mysuru. He is currently an Additional Professor with the Department of Mechanical and Industrial Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education. His research interests include machine learning, artificial intelligence, and explainable AI for engineering applications. He is a peer-reviewer in multiple engineering journals and has published many articles on artificial intelligence.

SHASHIKIRAN UMAKANTH received the M.B.B.S. and M.D. degrees in medicine. He is currently a Professor and a Medical Superintendent with the Dr. TMA Pai Hospital, Manipal Academy of Higher Education, Udupi, India. He has been a nodal head during the COVID-19 vaccine. He led the battle against the pandemic in Udupi. He has published many articles on medical AI applications. He is an FRCP Edin.