Realizing an Effective COVID-19 Diagnosis System Based on Machine Learning and IOT in Smart Hospital Environment

Karrar Hameed Abdulkareem, Mazin Abed Mohammed, Ahmad Salim, Muhammad Arif, Oana Geman, Deepak Gupta, Ashish Khanna

Abstract— **The aim of this study to propose a model based on Machine Learning (ML) and Internet of Things (IoT) to diagnose patients with COVID-19 in smart hospitals. In this sense, it was emphasized that by the representation for the role of ML models and IoT relevant technologies in smart hospital environment. The accuracy rate of diagnosis (classification) based on laboratory findings can be improved via light ML models. Three ML models, namely, Naive Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM), were trained and tested on the basis of laboratory datasets. Three main methodological scenarios of COVID-19 diagnoses, such as diagnoses based on original and normalized datasets and those based on feature selection, were presented. Compared with benchmark studies, our proposed SVM model obtained the most substantial diagnosis performance (up to 95%). The proposed model based on ML and IoT can be serve as a clinical decision support system. Furthermore, the outcomes could reduce the workload for doctors, tackle the issue of patient overcrowding, and reduce mortality rate during the COVID-19 pandemic.**

Index Terms— **COVID-19, Machine Learning, IOT, Smart Hospital Environment, laboratory findings, Naive Bayes, Random Forest, and Support vector machine.**

I. INTRODUCTION

THE services offered in hospitals are a great representation of healthcare. A wide range of specializations and services are offered to different groups of patients with various needs [1]. One of the problems being experienced in hospitals is overcrowding. This problem occurs when the number of patients is greater than that of available resources, including medical equipment, medical staff, or hospital care areas. When hospitals are overcrowded, the waiting time of patients increases, whereas the quality of service and the efficiency of healthcare providers decrease. The issue of overcrowding is a global issue [2] and a public health problem that affects individuals irrespective of one's social class [3]. The occurrence of this problem is attributed to certain factors, including the inflow and outflow of patients, patients' waiting time, and efficiency of healthcare providers, agglomerations, and availability of resources [4]. As a result of overcrowding, the quality of service decreases, the risk of mortality increases, and social discrimination prevails in the queues [5]. The authors in [5] found a strong relationship between overcrowding and mortality rate among patients. With the rate at which COVID-19 infection is spreading, an increase in all the aforementioned scenarios has been observed. Therefore, the increasing rate of COVID-19 poses a new challenge to the health sector alongside the expected provided service per day where rate of mortality is expected to be high and diagnosis process takes long time.

There is considerable research in the literature on the use of the Internet of Things (IoT) to improve health services especially during COVID-19 pandemic. For instance, during quarantine, it is helpful for a proper monitoring system. All high-risk patients are tracked easily using the internet-based network [6, 7]. Also, this technology is used for the biometric measurements in COVID-19 cases to verify blood pressure, heart beat and glucose level [8, 9]. The concept of Internet of Things (IoT) involves the using, processing, and storing information in the cloud; such information is made accessible and can be used independently by intelligent objects connected to the cloud via the Internet [10]. Through IoT, data can be shared and processed so that objects that are specifically designed for the improvement of life can be smartly integrated [1]. The typical representation of the application of IoT in service systems can be seen in smart hospitals that function on the basis of the IoT technology; the construction of such a smart hospital uses the vectors of different application services. This kind of hospital is regarded as a new type of hospital where different functions, including diagnosis, treatment, management and decision, are combined [11]. In addition, smart hospitals are designed by combining

Abdulkareem K. H., College of Agriculture, Al-Muthanna University, Samawah 66001, Iraq, khak9784@mu.edu.iq

Mazin Abed Mohammed, College of Computer Science and Information Technology, University of Anbar, 11, Ramadi, Anbar, Iraq; mazinalshujeary@uoanbar.edu.iq

Salim A., Middle Technical University, Baghdad, Iraq, ahmadsalim@mtu.edu.iq

Arif M., School of Computer Science, Guangzhou University, Guangzhou 510006, China; Email: arifmuhammad36@hotmail.com

Geman O., Department of Health and Human Development, Universitatea Stefan cel Mare din Suceava,, Suceava, 720229 Romania; E-Mail: oana.geman@usm.ro

Gupta D., Maharaja Agrasen Institute of Technology, Sector-22, Rohini, Delhi, India, email: deepakgupta@mait.ac.in

Khanna A., Maharaja Agrasen Institute of Technology, Sector-22, Rohini, Delhi, India, email: ashishkhanna@mait.ac.in

the concepts and ideas of intelligent, informative, and digital hospitals [12]. Integrated hospitals reflect the actual dynamic, elaborate, and specific description of hospitals. When a smart hospital is implemented, it may require the implementation of an application system based on a digital environment. With a smart hospital, relevant information about a service can be rapidly and accurately obtained by people. Therefore, this smart hospital system facilitates the informatization of diagnosis, scientific decision, and management standardization. Moreover, with the combination of application service in the system of smart hospitals, information can be acquired and shared in the hospital. The aim of this is to enhance the implementation of smart diagnosis, smart management, smart service, and smart treatment [12].

COVID-19 is often detected by using molecular approaches, such as quantitative real-time reverse transcription–polymerase chain reaction [13] and other methods, such as serologic tests [14] and viral throat swab testing [15]. However, previous studies have revealed that abnormalities pointing toward lung disease, including COVID-19, can be efficiently detected using chest radiographs (X-rays) [16] and chest computed tomography (CT) scans [17]. The severity of COVID-19 can be assessed by using X-ray tests and CT scans as main detection tools; these tools can also be used for monitoring the emergency case of infected patients and predicting the progression of COVID-19 [18]. However, such emergency situations are accompanied by time limitation; thus, the use of extant conventional manual diagnosis cannot be used [19]. Such procedures are prone to human error; thus, a specialist doctor is required during the testing, reading, and interpretation of the findings because errors can lead to inaccurate findings hence inappropriate treatment. With a spike in the transmission rates of COVID-19 worldwide, hospitals are filled with patients whose health conditions are improving or worsening [20]. Thus, the patient test must be conducted with a high level of efficiency and speed to save as many lives as possible [14]. Intelligent technologies are great assets that can assist in the effective diagnosis and classification of the severity of COVID-19 [16].

Different fields, particularly in the area of medical diagnosis and disease detection, are experiencing the increasing growth in the use of artificial intelligence (AI) [21]. In different fields, the use of AI has been extensively used because it facilitates the production of accurate detection results while reducing the workload of healthcare systems [22]. Additionally, AI can reduce the time required for decision-making during the process of detection when using traditional methods [23]. One of the main ways through which the prediction, prevention, and detection of future global health risks can be improved is by developing AI in a manner that allows the recognition of the risks of epidemic diseases [24]. A few researchers have introduced different kinds of AI classifiers that have been evaluated using real COVID-19 datasets with different targets and case studies [18]. Despite the benefits offered by AI techniques in terms of diagnosing and classifying COVID-19,

the selection of an AI technique that is appropriate for the production of accurate results remains a major challenge [25, 26]. The difficulty associated with the selection of the most appropriate technique for the diagnosis and classification of COVID-19 is attributed to the availability of a wide range of AI techniques [27].

The diagnosis of patients with COVID-19 on the basis of ML models and laboratory findings has found in the study [28]. Some authors have presented a remarkable contribution when they performed filtering and balancing for a dataset that contains 111 laboratory findings from 5644 various patients. They found that only 18 out 111 laboratory findings are important among 600 patients. This dataset has been tested with different deep learning models, and the best accuracy achieved is 92.3% by the CNN-LSTM hybrid model. Even with this remarkable diagnosis accuracy, accurate ML models are still needed, and the accuracy of diagnosis can still be improved. Deep learning models are based on several parameter settings and deep layers [29- 32]; thus, these models cannot be easily implemented in real-time applications because it needs numerous resources. In other words, these models are not lightweight architectures. Furthermore, the hybrid model (CNN-LSTM) is complex, and it hardly satisfies the resource requirements for this type of models in real-time environments. Moreover, the same study adopted several selected features based on the recommendation of other studies from the medical point of view; meanwhile, the authors have ignored the feature selection approach based on the requirements of ML models (technical point of view), especially because the volume of COVID-19 patient data is unpredictable and urgent medical involvement is needed.

Our study has the following objectives and contributions:
• To the best of our knowledge, this study is the first to use three light ML algorithms for COVID-19 prediction based on laboratory data.
• The roles of ML and IoT are highlighted in smart hospital environments to solve overcrowding problem during the COVID-19 pandemic.
• This study validates the RF, Bernoulli (NB), and SVM methods on COVID-19 diagnosis results based on the original, normalized datasets, and those based on selected laboratory findings using the brute-force feature selection technique.
• The accuracy of the COVID-19 diagnosis model is improved.

The rest of this paper is organized as follows: Section II presents the methodology, which is divided into two main stages; the first illustrates the roles of ML and IOT in smart hospital environments, and the second explains the steps of the COVID-19 diagnosis model within a target case study. In Section III, the results of 18 laboratory findings for 600 patients are used in the proposed COVID-19 prediction approach; three experimental results for the original and normalized datasets and that with feature selection are presented. Finally, Section IV presents the conclusions and recommendations for future work.

II. METHODOLOGY

The methodology of this study is divided into two stages. The first stage illustrates the role of ML and IOT in smart hospital environments. The second stage explains the steps of the COVID-19 diagnosis model within a target case study.

A. ML and IOT Model in Smart Hospital Environments

Smart hospitals, which are a novel type of hospitals, are based on the technology of ML and IoT built with the vector of various application service systems. These hospitals typically reflect two technologies that have been applied in special areas of healthcare. Smart hospitals are considered new because it integrates different functions, such as diagnosis, treatment, management, and decision-making. The evolution of ML and IoT in smart hospitals during the COVID-19 pandemic can be explained in three stages (see Fig 1).

First stage: Data generation and collection are conducted. The suspected patients are examined using different medical devices, such as PCR, CT scans, and X-rays. Then, the examination data are generated for further analysis. To capture the X-ray or CT scans, a lab technician performs these procedures remotely from a control room through a live streaming of the video images, which can be further processed. In this manner, the time required for the examination and production of base data of the COVID-19 case, is reduced. Moreover, with this procedure, less contact is needed. In other words, lab technicians do not need to move from door to door to conduct examinations.

Second stage: Numerous patients are investigated every single day during the COVID-19 pandemic; thus, a huge volume of patient data is generated every minute. The problem of overcrowding occurs when the number of patients waiting to be attended to supersedes that of the available of resources (medical equipment, medical practitioners, or hospital care areas). Furthermore, the historical data for each COVID-19 patient are valuable for further analysis. With limited resource capacity in smart hospitals, a practicable solution is required. However, the collected data from the first stage are transferred through the Internet to the IoT cloud where services facilitate the instantaneous and on-demand delivery of computing infrastructure, databases, storage, and applications needed for the processing and analysis of COVID-19 patient data generated from hundreds of cases (shown in Fig 1). In this stage, patient data are stored for further analysis. Then, when a request for COVID-19 patient diagnosis is requested, data are transferred from the cloud storage to a ML diagnosis model. However, as mentioned earlier, numerous patients must be diagnosed each day during the COVID-19 pandemic. Furthermore, the overcrowding of patients, along with a shortage of doctors for patient diagnosis, poses a huge challenge; thus, a feasible solution is needed. Therefore, in stage one of our proposed methodology, an ML model is adopted to serve as a decision support system for the diagnosis of patients with COVID-19. In this manner, the overcrowding issue and doctors' workload can be reduced. Usually, the laboratory findings transferred for IoT cloud are fed into ML models to diagnose if the patient is positive or negative for COVID-19. The implementation of the diagnosis model is conducted in a workstation, where the ML model can be trained offline. Through software integrated into the system, the training process can be performed automatically with new data to facilitate the updating of the cloud and improve diagnosis capabilities. Further details of proposed ML model are highlighted in Sub-section B.

Third stage: After the diagnosis results are obtained by the ML model, patient health reports are generated and then sent to a decision-making unit where the next action could either be patient quarantine or patient to be brought to the emergency room, and so on.

B. COVID-19 Diagnosis Model

This section explains the proposed ML diagnosis model, which is composed of three main phases as shown in Fig 2. **Phase one**: Three main sub-phases are conducted in this phase.

First, the base dataset for the experiment was extracted from [28]. This study used a laboratory dataset of patients with COVID-19 in the Israelita Albert Einstein Hospital in Sao Paulo, Brazil. The patient samples were collected to identify who were infected by COVID-19 in the beginning of year 2020. The laboratory dataset contains information on 600 patients with 18 laboratory findings. In this dataset, 520 had no findings, and 80 were patients with COVID-19.

Second, the relevant information extracted from raw data is critical in COVID-19 classification due to its direct influence on classification performance. The main input for the classification of COVID-19 cases are values with each laboratory finding for the mentioned dataset. In this stage, 18 laboratory findings (i.e., red blood cells, hemoglobin, platelets, hematocrit, aspartate transaminase, lymphocytes, monocytes, sodium, urea, basophils, creatinine, serum glucose, alanine transaminase, leukocytes, potassium, eosinophils, C reactive protein, and neutrophils) are included in the feature extraction process.

Third, the diagnosis process for COVID-19 patient data was associated with a classification approach. The diagnosis stage was conducted on the basis of the defined dataset, extracted features, and the implementation of different ML models. Python was used for all classification tasks during the research period. The final output of the diagnosis for COVID-19 was generated in this sub-phase where binary classifcation was conducted. In other words, two cases are the ML models need to predicate namely, normal and COVID-19 cases.

Phase two:

In this phase, almost the same processes were applied, except for the data normalization process. The feature extraction and diagnosis processes were conducted on the basis of the normalized dataset. However, attempts were made by several ML algorithms to identify trends in the data by comparing features of data points. However, an issue emerged due to the great variations in the scales of the features. Normalization is specifically aimed at making every data point have the same scale because each feature has equal relevance. One of the commonly used methods of data normalization is min-max normalization, which involves the transformation of the minimum value of each feature into 0, maximum value

into 1, and every other value into a decimal between 0 and 1.

The dataset had a large discrepancy in the values in one column or between the values in all the columns because of the existence of some negative values and some positive values. Thus, normalization was required to solve the problem of differences between the values. The formula for min-max normalization is as follows:

$$v' = \frac{v - min}{max - min} \tag{1}$$

Phase three:

On the basis of the normalized dataset, three main processes were conducted, with only one process different from the previous phases, that is, feature selection was applied. The selection of laboratory features was performed on the basis of the brute-force feature selection method, which extensively evaluates all candidate combinations of the input features and then finds the most appropriate subset. Apparently, this extensive search incurs high computation cost and a relative risk of overfitting. Therefore, greedy methods should be considered as an alternative for the brute-force method of feature selection. In this study, real-world datasets obtained from more than 10 domains were used to compare the different algorithms in terms of efficiency and accuracy.

#	Algorithm 1: Pseudocode for the feature selection method
01	Start brute-force feature selection;
02	Set and of split the data set to of 80% of the data for training set and 20% for testing set.
03	Suppose I has different feature sizes i = (0 , 1 , … , n)
04	a: Suppose feature selection of I = best set size I, by calculated as minimizer of the best
05	Split data in the training process.
06	b: Suppose test feature = feature selection set of i;
	Choose the best feature set as feature selection i that minimizes the test set score

To find the optimal model for the diagnosis of patients with COVID-19, the performance of all ML diagnosis models was measured on the basis of an evaluation experiment within each of the three phases. The evaluation experiment was conducted on the basis of the following measurements:

$$Accuracy = ((TP + TN)/(TP + TN + FP + FN)) \times 100, \tag{2}$$
$$Precision = TP/(TP + FP), \tag{3}$$
$$Recall = TP/(TP + FN), \tag{4}$$
$$F1 = (Precision * Recall)/(Precision + Recall \tag{5}$$

Where TP stands for true positive, TN denotes true negative, FP represents false positive, and FN stands for false negative cases. Based on the evaluation experiment results for the different ML models, the optimal model was chosen in accordance with the highest measurement values in the three phases.

III. RESULTS AND DISCUSSION

In this research, 18 laboratory findings for 600 patients were used in the proposed COVID-19 prediction method. The data of 600 patients and tests were obtained from a laboratory dataset. Three diverse ML methods, namely, RF classifier, Bernoulli NB, and SVM, were improved and used as ML classifiers. In this section, the results of the proposed model are analyzed and discussed in four Sub-sections: COVID-19 diagnosis results based on the (A) original and (B) normalized datasets, (C) those based on selected laboratory findings, and (D) those against benchmark studies.

A. COVID-19 Diagnosis Results Based on the Original Dataset

In this research, we developed a fully automated approach for COVID-19 prediction using ML methods and medical IoT. Three ML methods (i.e., RF, SVM, and Bernoulli NB) were used on the original dataset. These methods were evaluated in terms of recall, precision, accuracy, F1 score, and AUC. The evaluation results of the methods with an 80–20 train–test split technique are shown in Table 1.

TABLE 1: EVALUATION RESULTS OF THE ML METHODS WITH AN 80–20 TRAIN–TEST SPLIT TECHNIQUE ON THE ORIGINAL DATASET

Classifiers	F1	Precision	Recall	Accuracy	AUC
RF	89	88	90	90	0.78
Bernoulli NB	87	85	89	89.16	0.82
SVM	92	94	93	93.33	0.88

In this study, we observed that the best prediction methods are measured by AUC metric is 0.88 for SVM classifier as best COVID-19 predication model in the predictive performance. COVID-19 prediction based on laboratory findings is a difficult task and undergoes a challenging process because sample collection requires a long time and a complex process. Therefore, the best laboratory findings identification outcomes achieved with great evaluation measurements of F1-score of 92%, accuracy of 93.33%, precision of 94%, and recall of 93%, respectively for SVM classifier. The SVM classifier achieves good results because it is widely used in disease prediction and data sequences, especially when data include a time series. The results of the model evaluation are shown in Fig 3.

B. COVID-19 Diagnosis Results Based on the Normalized Dataset

The number of COVID-19 cases increases every day; thus, it takes time to understand laboratory findings. Consequently, impediments in terms of findings and treatments also increase. Given these restrictions and challenges, the need for clinical support approaches with prediction systems has emerged. Prediction systems may possibly ease the strain on medical centers and hospitals by distinguishing COVID-19 cases. To achieve good decision-making and accurate diagnoses, we used min-max approach to normalize and filter the dataset. Numerous ML techniques have been used to discover patterns within data by comparing data point features. Thus, a challenge for feature extraction is the different scales of features. The objective of normalization is to make each data point have the same scale for all the features and make these

features similarly critical. Min-max normalization is one of the most popular methods used to normalize data. In this method, the minimum value of every feature is transformed to 0, the maximum value is transformed to 1, and other values are transformed to a specific ratio between 0 and 1.

We repeated the same experiment using three ML methods (RF, SVM, and Bernoulli NB) on the normalized dataset and evaluated these ML methods in terms of recall, precision, accuracy, F1 scores, and AUC. According to [28], the clinical predictive algorithms can achieve better predication performance with 80–20 splitting approach. The evaluation results of the methods with an 80–20 train–test split technique is shown in Table 2.

TABLE 2: EVALUATION RESULTS OF THE ML METHODS WITH AN 80–20 TRAIN–TEST SPLIT TECHNIQUE ON THE NORMALIZED DATASET

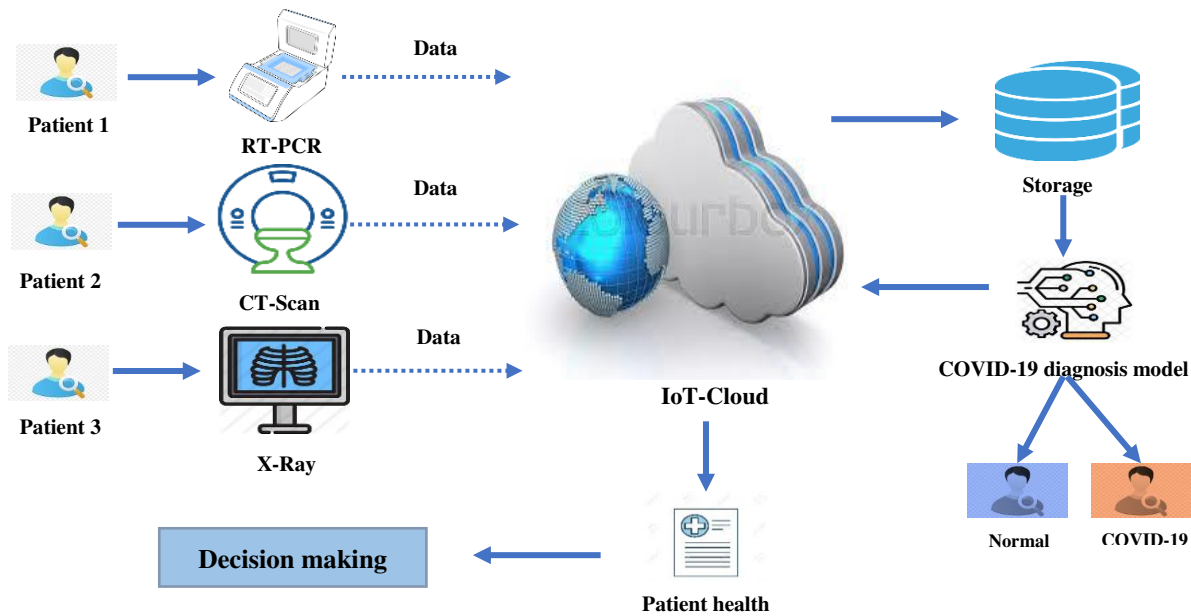| Classifiers	F1	Precision	Recall	Accuracy	AUC
RF	92	93	93	93.33	0.89
Bernoulli NB	91	92	91	90.83	0.81
SVM	92	94	94	94.16	0.85

Fig. 1. Role of IOT and ML in smart hospitals during the COVID-19 pandemic.
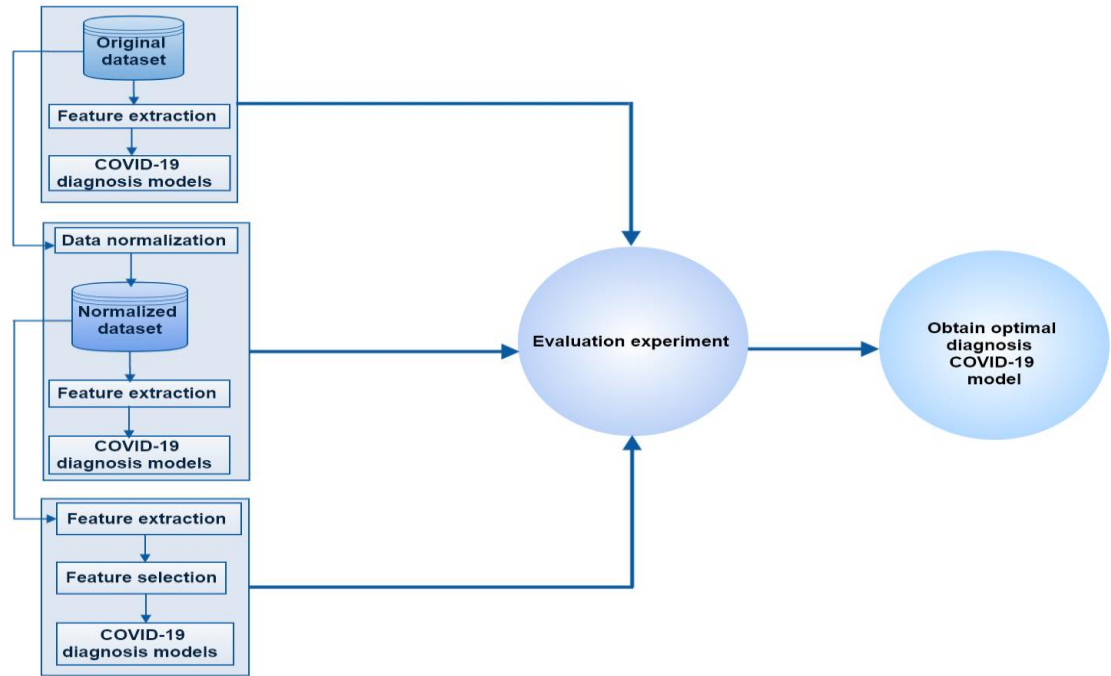
Fig. 2. Methodological steps for the proposed COVID-19 diagnosis model.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3050775, IEEE Internet of Things Journal
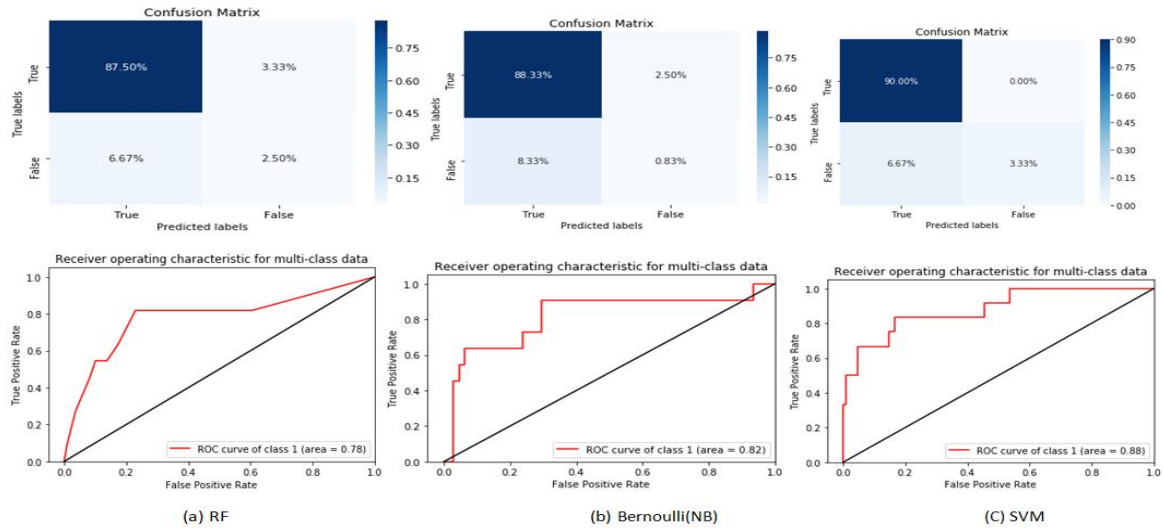
6

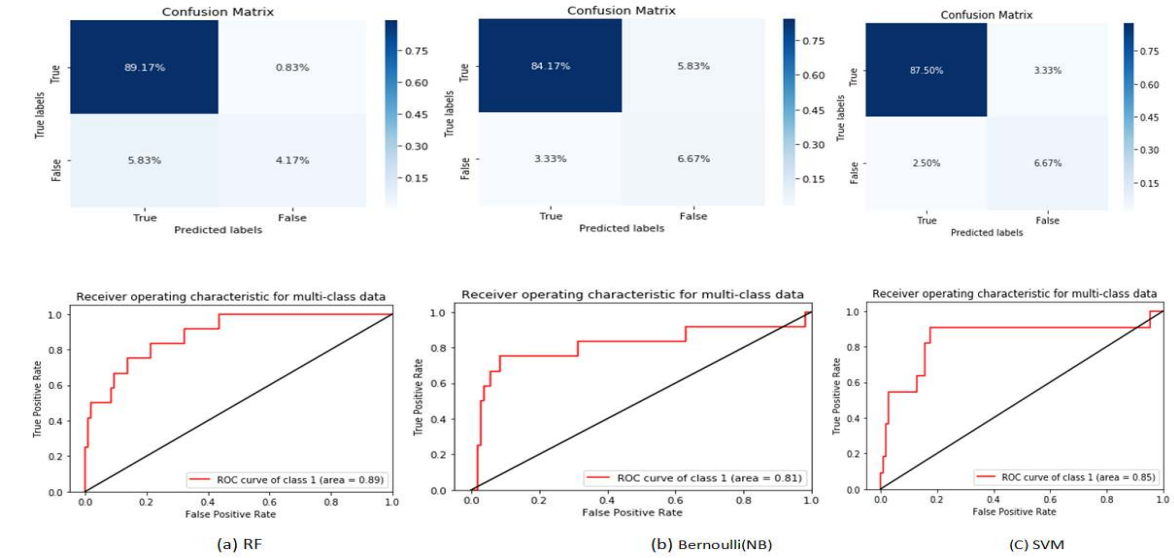Fig. 3. The confusion matrix and ROC curve for three models based on original dataset.

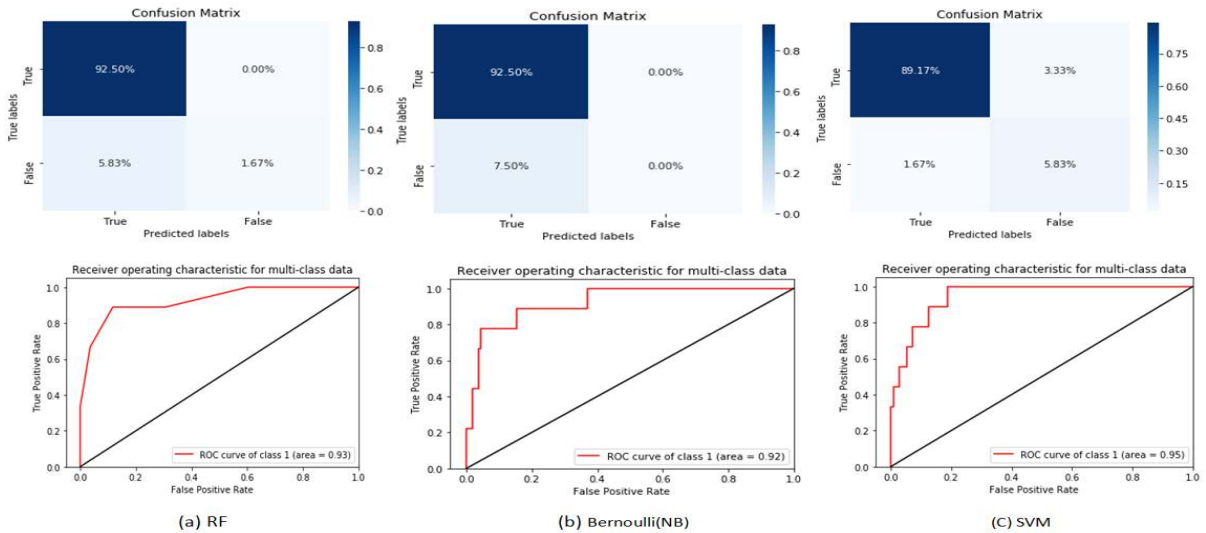Fig. 4. The confusion matrix and ROC curve for three models based on normalized dataset.

Fig. 5. The confusion matrix and ROC curve for three models with constraints of the normalized dataset and feature selection.

C. COVID-19 Diagnosis Results Based on Selected Laboratory Findings

To provide an improved COVID-19 identification model using improved and different ML methods. To the best of our knowledge, no research has used different ML techniques with IoT for COVID-19 prediction based on laboratory findings. The present study may encourage researchers to validate the methods by using various laboratory data. To improve accuracy and identify the best model, we used the brute-force technique as the feature selection process to discover the most important laboratory features. Feature selection is important for ML methods because insignificant features influence the prediction execution of ML methods. Feature selection enhances the precision of prediction and decreases the execution time of ML methods. The brute-force feature selection technique aims to assess all conceivable combinations of the input features thoroughly; afterward, the most excellent subset is discovered. Clearly, the comprehensive search's computational fetched is restrictively tall, with impressive threat of overfitting. Consequently, individuals resort to greedy techniques, such as forward determination.

We repeated the third experiment on the normalized dataset and used the brute-force feature selection technique with three ML methods. The evaluation results of the ML methods with an 80–20 train–test split technique on the normalized dataset, along with brute-force feature selection, are shown in Table 3.

TABLE 3: EVALUATION RESULTS OF ML METHODS WITH AN 80–20 TRAIN–TEST SPLIT TECHNIQUE ON THE NORMALIZED DATASET, ALONG WITH BRUTE-FORCE FEATURE SELECTION

Classifiers	F1	Precision	Recall	Accuracy	AUC
RF	93	95	94	94.16	0.93
Bernoulli NB	89	86	93	92.5	0.89
SVM	94	95	95	95	0.95

The importance of feature selection has been shown on the COVID-19 laboratory findings when we compared the difference among the accuracies of the ML methods. From 18 clinical features, we selected the best 15 features and dropped three features (i.e., monocytes, sodium, and alanine transaminase). Which is almost consistent with selected features based on the medical point of view. However, the best laboratory findings identification outcomes achieved with great evaluation measurements of F1-score of 94%, accuracy of 95%, precision of 95%, and recall of 95%, respectively for SVM classifier. The results of the models' evaluation are shown in Fig 5.

D. COVID-19 Diagnosis against Benchmark Studies

Benchmarking is an essential step that must be used in medical processing and diagnosis research to determine the efficiency and reliability of developed approaches. It is usually conducted by using a standard dataset or approaches that have been used for the same problem domain or application. Moreover, benchmarking is achieved utilizing the best methods for COVID-19 laboratory findings based on ML approaches and feature selection methods existed in the literature. Table 4 presents the benchmark studies used in the present work.

TABLE 4: PERFORMANCE RESULTS OF THE PROPOSED METHOD VERSUS THE STATE-OF-THE-ART TECHNIQUES FOR COVID-19 LABORATORY FINDINGS

Work	Method	Accuracy	F1	Precision	Recall	AUC
[33]	SVM, RF	-	0.72	-	-	0.87
[34]	XGB	-	-	-	-	0.66
[28]	CNNLSTM	92.30	0.93	-	-	0.90
Our Work	SVM	**95**	**94**	95	95	**95**

As shown in Table 4, different studies have been used and compared on the same dataset with different traditional ML and deep learning methods. Our SVM model achieved better results in accuracy and other measures than other studies. Based on the findings, our study is the first to use different measurements and achieve good results. Our proposed SVM classifier is considered the best method, with an accuracy of 95%, F1 score of 94%, precision of 95%, recall of 95%, and AUC of 95%.

The main limitation associated with our study is data size. The data used in this study comprise information on 600 patients, with some laboratory findings for some patients not measured. Thus, our prediction system has a success range of 90%–95%, depending on the dataset used, type of the data used (original or normalized), and type of feature selection method. Additionally, the laboratory data used in our study were imbalanced; therefore, we balanced this dataset by withdrawing some features and using a normalization method. The execution of these methods can be improved with a larger dataset. Additional experiments and research must be conducted with laboratory findings from other areas to confirm these outcomes.

IV. CONCLUSION

During the COVID-19 pandemic, numerous patients are expected to be diagnosed, treated, and monitored, thus bringing a huge burden to medical organizations. Adopting additional automated services can reduce the workload for doctors, overcrowding, and mortality rate. Moreover, in smart hospitals, ML approaches and IoT paradigms can serve as clinical decision support systems for handling issues related to the COVID-19 pandemic. To the best of our knowledge, this study is the first to propose a ML and IoT model based on laboratory findings in smart hospital environments. Different ML approaches were implemented, aiming to improve the accuracy of diagnosis for COVID-19 cases. The steps and representation for the proposed model in smart hospitals were described and explained. On the basis of a laboratory dataset, three different experiments were conducted to find the most optimal diagnosis results among the selected ML models. Then, the best ML approach was compared with benchmark studies that have adopted the same laboratory dataset. The results of this study confirm that (1) based on the diagnosis results in the original COVID-19 laboratory dataset, SVM outperforms ML models in all evaluation metrics; (2) based on the normalized dataset diagnosis results, SVM outperforms

other ML models in all metrics, except for AUC where the highest value was obtained by the RF model (89%), in the same time each of SVM and RF have scored 92% as F1-score value; (3) based on the feature selection approach results, only 15 out of 18 features are effective for the diagnosis of COVID-19 cases, thus reducing the computational load and diagnosis time for ML models. Furthermore, SVM outperforms other models in all evaluation metrics, except for precision where SVM shares an equal value (95%) with RF. (4) Finally, compared with other studies that have adopted the same laboratory dataset, the proposed SVM shows a remarkable improvement in terms of diagnosis accuracy (up to 2.7%); the model scored 95% in accuracy, 94% in F1, and 95% in precision, recall, and AUC. The results of this study are limited within only three ML models. However, in the future authors plan to investigate more laboratory datasets, ML models, and different feature selection methods.

Author Contributions: Conceptualization, O.G., K. H, M.M.; methodology, K.H, M.M; software, M.M. and A.S, K.H, formal analysis, K.H, M.M, A.S, M.A, O.G, D.G.; investigations K.H., O.G., A.A.; resources, K.H. data analysis K.H., O.G.,K.A.; writing—original draft preparation, K.H, M.M, A.S,; writing—review and editing, K.H, M.M, A.S, M.A, O.G, D.G, and K. A. .All authors have read and agreed to the published version of the manuscript.

Ethical Statement: This study used a laboratory dataset of patients with COVID-19 in the Israelita Albert Einstein Hospital in Sao Paulo, Brazil. This dataset available online and anyone can be used.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

[1] C. Cáceres, J. M. Rosário, and D. Amaya, "Proposal of a smart hospital based on Internet of Things (IoT) concept," in Sipaim–Miccai Biomedical Workshop, 2018, pp. 93-104: Springer.
[2] K. Jensen, "Emergency department crowding: the nature of the problem and why it matters," in Patient Flow: Springer, 2013, pp. 97-105.
[3] J. M. Pines et al., "International perspectives on emergency department crowding," vol. 18, no. 12, pp. 1358-1370, 2011.
[4] R. J. Bittencourt and V. A. J. C. d. S. P. Hortale, "Intervenções para solucionar a superlotação nos serviços de emergência hospitalar: uma revisão sistemática," vol. 25, no. 7, pp. 1439-1454, 2009.
[5] M. McHugh, "The consequences of emergency department crowding and delays for patients," in Patient Flow: Springer, 2013, pp. 107-127.
[6] Singh, R.P., Javaid, M., Haleem, A. and Suman, R., 2020. Internet of things (IoT) applications to fight against COVID-19 pandemic. Diabetes & Metabolic Syndrome: Clinical Research & Reviews.
[7] Mohammed, M.N., Syamsudin, H., Al-Zubaidi, S., AKS, R.R. and Yusuf, E., 2020. Novel COVID-19 detection and diagnosis system using IOT based smart helmet. International Journal of Psychosocial Rehabilitation, 24(7).
[8] Raju Vaishya, Mohd Javaid, Ibrahim Haleem Khan, Abid Haleem, Artificial Intelligence (AI) applications for COVID-19 pandemic, Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 14(4), 2020.
[9] Al-Waisy, A.S., Al-Fahdawi, S., Mohammed, M.A. et al. COVID-CheXNet: hybrid deep learning framework for identifying COVID-19 virus in chest X-rays images. Soft Comput (2020). https://doi.org/10.1007/s00500-020-05424-3
[10] D. Bradley, D. Russell, I. Ferguson, J. Isaacs, A. MacLeod, and R. J. M. White, "The Internet of Things–The future or the end of mechatronics," vol. 27, pp. 57-74, 2015.
[11] K. Dhariwal and A. J. I. R. J. E. T. Mehta, "Architecture and plan of smart hospital based on Internet of Things (IOT)," vol. 4, no. 4, pp. 1976-1980, 2017.
[12] M. J. I. B. JinJun and C. Information, "A brief talk on the problem in integration of hospital intelligence and information and developing directionof intelligence [J]," vol. 158, pp. 94-96, 2010.
[13] Y. Fang et al., "Sensitivity of chest CT for COVID-19: comparison to RT-PCR," p. 200432, 2020.
[14] H. Zeng et al., "Antibodies in infants born to mothers with COVID-19 pneumonia," vol. 323, no. 18, pp. 1848-1849, 2020.
[15] D. A. J. A. o. p. Schwartz and l. medicine, "An analysis of 38 pregnant women with COVID-19, their newborn infants, and maternal-fetal transmission of SARS-CoV-2: maternal coronavirus infections and pregnancy outcomes," vol. 144, no. 7, pp. 799-805, 2020.
[16] T. Ozturk et al., "Automated detection of COVID-19 cases using deep neural networks with X-ray images," p. 103792, 2020.
[17] M. Li et al., "Coronavirus disease (COVID-19): spectrum of CT findings and temporal progression of the disease," 2020.
[18] F. Shi et al., "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19," 2020.
[19] H. S. Maghdid, A. T. Asaad, K. Z. Ghafoor, A. S. Sadiq, and M. K. J. a. p. a. Khan, "Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms," 2020.
[20] W. H. Organization, "Coronavirus disease 2019 (COVID-19): situation report, 72," 2020.
[21] D. D. Miller and E. W. J. T. A. j. o. m. Brown, "Artificial intelligence in medical practice: the question to the answer?," vol. 131, no. 2, pp. 129-133, 2018.
[22] K.-H. Yu, A. L. Beam, and I. S. J. N. b. e. Kohane, "Artificial intelligence in healthcare," vol. 2, no. 10, pp. 719-731, 2018.
[23] A. Albahri and R. A. J. J. o. M. S. Hamid, "Role of biological Data Mining and Machine Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19): A Systematic Review," vol. 44, no. 7, 2020.
[24] Z. Yang et al., "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions," vol. 12, no. 3, p. 165, 2020.
[25] M. Alsalem et al., "Systematic review of an automated multiclass detection and classification system for acute Leukaemia in terms of evaluation and benchmarking, open challenges, issues and methodological aspects," vol. 42, no. 11, p. 204, 2018.
[26] M. Alsalem et al., "Multiclass benchmarking framework for automated acute Leukaemia detection and classification based on BWM and group-VIKOR," Journal of medical systems, vol. 43, no. 7, p. 212, 2019.
[27] O. Albahri et al., "Systematic review of artificial intelligence techniques in the detection and classification of COVID-19 medical images in terms of evaluation and benchmarking: Taxonomy analysis, challenges, future solutions and methodological aspects," 2020.
[28] T. B. Alakus, I. J. C. Turkoglu, Solitons, and Fractals, "Comparison of deep learning approaches to predict COVID-19 infection," p. 110120, 2020.
[29] M. A. Mohammed et al., "Benchmarking Methodology for Selection of Optimal COVID-19 Diagnostic Model Based on Entropy and TOPSIS Methods," in IEEE Access, vol. 8, pp. 99115-99131, 2020, doi: 10.1109/ACCESS.2020.2995597.
[30] Mohammed, M.A., Abdulkareem, K.H., Mostafa, S.A., Ghani, M.K.A., Maashi, M.S., Garcia-Zapirain, B., Oleagordia, I., Alhakami, H. and AL-Dhief, F.T., 2020. Voice Pathology Detection and Classification Using Convolutional Neural Network Model. Applied Sciences, 10(11), p.3723.
[31] Abd Ghani, M.K., Mohammed, M.A., Arunkumar, N., Mostafa, S.A., Ibrahim, D.A., Abdullah, M.K., Jaber, M.M., Abdulhay, E., Ramirez-Gonzalez, G. and Burhanuddin, M.A., 2020. Decision-level fusion scheme for nasopharyngeal carcinoma identification using machine learning techniques. Neural Computing and Applications, 32(3), pp.625-638.
[32] Mutlag, A.A., Khanapi Abd Ghani, M., Mohammed, M.A., Maashi, M.S., Mohd, O., Mostafa, S.A., Abdulkareem, K.H., Marques, G. and de la Torre Díez, I., 2020. MAFC: Multi-Agent Fog Computing Model for Healthcare Critical Tasks Management. Sensors, 20(7), p.1853.
[33] de Moraes Batista, A.F., Miraglia, J.L., Donato, T.H.R. and Chiavegatto Filho, A.D.P., 2020. COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. doi: 10.1101/2020.04.04.20052092
[34] Schwab, P., Schütte, A.D., Dietz, B. and Bauer, S., 2020. predCOVID-19: A Systematic Study of Clinical Predictive Models for Coronavirus Disease 2019. arXiv preprint arXiv:2005.08302.