Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Real-time data of COVID-19 detection with IoT sensor tracking using artificial neural network☆

Roaa Mohammedqasem*, Hayder Mohammedqasim, Oguz Ata

Department Of ECE, Institute Of Science, Altinbas University, Istanbul, Turkey

ARTICLE INFO

Keywords:
COVID-19
Synthetic minority oversampling technique
Recursive feature elimination
Imbalanced Dataset
Internet of Things

ABSTRACT

The coronavirus pandemic has affected people all over the world and posed a great challenge to international health systems. To aid early detection of coronavirus disease-2019 (COVID-19), this study proposes a real-time detection system based on the Internet of Things framework. The system collects real-time data from users to determine potential coronavirus cases, analyses treatment responses for people who have been treated, and accurately collects and analyses the datasets. Artificial intelligence-based algorithms are an alternative decision-making solution to extract valuable information from clinical data. This study develops a deep learning optimisation system that can work with imbalanced datasets to improve the classification of patients. A synthetic minority oversampling technique is applied to solve the problem of imbalance, and a recursive feature elimination algorithm is used to determine the most effective features. After data balance and extraction of features, the data are split into training and testing sets for validating all models. The experimental predictive results indicate good stability and compatibility of the models with the data, providing maximum accuracy of 98% and precision of 97%. Finally, the developed models are demonstrated to handle data bias and achieve high classification accuracy for patients with COVID-19. The findings of this study may be useful for healthcare organisations to properly prioritise assets.

1. Introduction

The Internet of Things (IoT) connects tangible objects to the Internet and enables them to send and receive data. The concept of the IoT is influenced by sensors, machine learning (ML), real-time analytics, and embedded systems. In December 2019, the coronavirus disease (COVID-19) was first reported in Wuhan, China, that rapidly spread around the world. The World Health Organisation [1] declared the COVID-19 outbreak a pandemic and recommended various equipment and mechanisms to rapidly detect cases with the maximum threat of illness and mortality. The coronavirus pandemic has posed a challenge to international health systems, leading to drastic increased requirement of testing and hospitalisation for diagnosis and treatment of affected patients, given the uncontrolled

Abbreviations: DL, Deep learning; ML, Machine learning; ANN, Artificial Neural Network; CNN, Convolutional Neural Network; RNN, Recurrent Neural Network; IoT, Internet of Things; AUC, Area Under Curve; SMOTE, Synthetic Minority Oversampling Technique; RFE, Recursive Feature Elimination; COVID-19, Coronavirus disease.

☆ This paper is for special section VSI-iot. Reviews were processed by Guest Editor Prof. Mustafa Matalgah and recommended for publication.

* Corresponding author.

E-mail addresses: Roaa.mohammedqasem@gmail.com (R. Mohammedqasem), haydermohammedqasim@gmail.com (H. Mohammedqasim), oguzata@gmail.com (O. Ata).

https://doi.org/10.1016/j.compeleceng.2022.107971
Received 24 December 2021; Received in revised form 26 March 2022; Accepted 28 March 2022
Available online 6 April 2022
0045-7906/© 2022 Elsevier Ltd. All rights reserved.
emergence of different variants of SARS-CoV2. Although COVID-19 affects humans in various ways, more than 80% of the cases with mild-to-moderate disease improve without hospitalisation. The most common signs and symptoms of this disease are fever, dry cough, and tiredness, which appear with mild intensity at the onset in all infected people. Moreover, other signs such as chest pain and fatigue can appear in some patients [2]. Artificial intelligence (AI) and deep learning (DL) algorithms, which can be considered a useful technology for improving common diagnostic methods such as chest X-ray, computed tomography (CT), and reverse transcription polymerase chain reaction (RT-PCR), have been applied to accurately diagnose COVID-19. As an essential approach to rapidly identify infection, development of technologies such as non-human thermal screening using drones is desired [3].

The main advantage of AI-based platforms is that they accelerate the diagnosis and treatment of COVID-19 [4], thus showing great potential to drastically improve and promote health research. Among the various applications of AI to solve real-time problems [5], its application to effectively manage health crises can help improve the accuracy and speed of case identification through data retrieval. The applications of ML methods and statistical models to perform various tasks without explicit commands have been described in a systematic review [6], and ML techniques are currently being used for predictions all over the world because of their accuracy. However, the ML approach has some limitation due to factors such as the low quality of online databases. One of the problems with model learning or selecting the optimal ML model for forecasting, for example, is determining the appropriate parameters [7]. To achieve predictions based on the available data, researchers use ML models that best fit the dataset.

Previous studies have reported that the effectiveness of medical data models depends primarily on increasing recognition of classifiers and retraining of reference models [8,9]. The problem of imbalance in medical datasets, which biases the model towards the group with the largest sample size in the training process, remains to be addressed. Furthermore, feature selection is an important issue and includes selecting a subset of elements to represent data classes. The feature selection is the basis of two important aspects: First, the function must filter out noise and eliminate redundant functionality, which can lead to a significant loss of detection accuracy. In this study, a grid-search algorithm is evaluated and validated by focusing on the feature selection process by using an advanced system rather than a traditional approach. The findings of this study can be used to assess the effectiveness of a DL model in improving imbalanced datasets and achieving accurate detection systems.

In this study, a new hyper ML method is proposed to manage highly imbalanced datasets and contradicting features. In an imbalanced dataset, the number of samples in one class is extremely lower than that in other classes; the class containing the large number of samples is called the major class, whereas that with a few samples is called the minor class. For diagnosing COVID-19 infection, ML techniques have proven to be useful, owing to a dramatic increase in the percentage of infected people and lack of experience about this disease [10]. As one of the powerful ML techniques, the synthetic minority oversampling technique (SMOTE) can solve the imbalanced dataset problem by generating new samples for the minor class in the dataset by using the k-nearest neighbour (KNN) algorithm and subsequently providing nearly equal classes. However, in this technique, highly relevant features are not effectively chosen from the original dataset, which leads to poor classification accuracy. This problem can be addressed using Recursive feature elimination (RFE), which is an effective technique for removing less-important features from the dataset. In this research, we develop and implement different ML systems for detecting COVID-19 infection by using the SMOTE and RFE on the basis of an optimised DL model.

The contributions of this paper are as follows: The hyperparameters for each DL model are optimised by grid search. Using a new hyper preprocessing approach to balance the data samples in each class and RFE, the processing time is reduced by removing contradictions from the dataset. Based on the proposed approach, the DL models using the COVID-19 dataset can accurately predict clinical outcomes. To our knowledge, a few studies have discussed COVID-19 infection with laboratory findings. Hence, this work may motivate researchers to test the models by using diverse laboratory data.

The remainder of the paper is organised as follows. Section 2 briefly presents the latest studies related to this work. Section 3 presents the methodology of the proposed ML system. Section 4 presents the results, including Fig.s and tables, of the performance evaluation experiments comparing the proposed method and previous methods. Section 5 presents the limitations of our research. Finally in Section 6, we present conclusions and future recommendations.

2. Related Work

Recently, extensive research has been conducted on COVID-19, with a focus on the origin, genetic structure, and distribution of the virus [11]. Prediction of the clinical outcome for patients affected by COVID-19 is, however, difficult because of the various confounding factors that affect patient classification, such as imbalanced datasets, nonsignificant features, and training time. Addressing these problems is crucial to help health organisations better understand and define medical outcome.

Early detection of COVID-19 is crucial to minimise transmission and improve patient care. Separating or isolating healthy persons from the infected or suspected COVID-19 carriers has proven to be the most effective measure to avoid COVID-19 transmission [12]. Radiology and imaging, particularly chest CT scans, are useful techniques for determining the COVID-19 stages and potential risks to the lungs of the patients [9]. META-COVID19 is a novel metaheuristic framework that employs generalised Boltzmann distribution and a Jacobi polynomial family to dynamically describe COVID-19 spread without any prior data or help of a human expert [13]. This method adopts the daily record of incidents as the input and outputs a polynomial mathematical model. The framework requires only one parameter, namely the number of Jacobi polynomials, to evaluate the entire iterative procedure. Finally, a theoretical mathematical study was used to establish the applicability of META-COVID19 models to comprehend features of COVID-19’s spread through time. Four ML classifiers as well as the SMOTE were applied. The Shapley Additive Explanations (SHAP) approach was used to determine the weightage of each feature; it was revealed that eosinophils, leukocytes, and platelets were the most important blood parameters for distinguishing COVID-19 infection in the samples. These classifiers can be used in combination with RT-PCR testing to
improve sensitivity as well as in emergency situations such as a pandemic outbreak caused by new viral strains. The promising findings of this study suggest that an automated framework might help doctors and medical workers diagnose and screen patients [14]. Next, ML classification methods were used to forecast the severity of COVID-19 infection [15]. In these methods, multilayered perception (MLP), XGBoost, and logistic regression categorised patients with 91% accuracy. The study also described possible fact-analysis programmes on various key aspects of COVID-19.

In another study [16], 11 clinical features were examined, and six different models were implemented: a KNN model, logistic regression model, two decision trees, a support vector machine (SVM) and a random forest classifier. The classifier worked best with the SVM classifier, providing 80% accuracy. In [17], the dataset collected from the Israel Albert Einstein Hospital in Sao Paulo, Brazil, which is the same as our dataset, was used to develop various DL applications to predict the diagnosis of COVID-19; among the compared models, the CNN-LSTM-hyper model performed the best with an accuracy of 92% and area under the curve (AUC) of 90%.

In another study, two ML models were used in the context of text analysis, and their effect on the classification of the length of tweets related to coronavirus was examined. The models achieved 91% accuracy for short tweets, whereas the Naive Bayesian approach and logistic regression approach provided conservative accuracy of up to 74% for shorter tweets [18]. Furthermore, an expert model mainly based on the artificial neural network (ANN) and deep extreme learning machine (DELM) was developed [19]. This model showed high potential for predicting coronavirus outbreaks in exclusive areas and achieved great strides. Furthermore, different numbers of hidden neurons were described, and multiple activation options and functions were used to accurately correlate various DELM parameters to achieve the optimal shape.

In the aforementioned studies that have used medical datasets, the majority of categories in the data were not improved, which leads to bias of the models towards the groups with the largest number of samples in the training process. Therefore, the present study aims to address this issue by developing a system that provides relatively higher prediction accuracy than previous systems.

3. Methodology

In this section, we describe the structure of our methodology used to solve the problems in medical datasets. The main objective of the proposed methodology is to create an efficient ML system that can enhance the prediction of an imbalanced medical dataset and other datasets.

Our methodology involves five steps: (1) preprocessing of the COVID-19 dataset, (2) splitting the dataset into training and testing sets, (3) establishing the theoretical frameworks of ML methods, (4) determining the optimal hyperparameters for the ML models, and (5) evaluating the ML models on various classification matrices. The stages of the methodology are outlined in Fig. 1 and described in detail in the following subsections.

3.1. Data Description

The dataset consists of several laboratory results that were obtained from the Albert Einstein Hospital located in Brazil, provided from Kaggle. Samples of patients with COVID-19 were collected during the primary months of 2020. The dataset consists of 111 laboratory results from 5,644 patients, without records on patients’ sex.

![Fig. 1. Machine learning flowchart for COVID-19 models.](image-url)
3.2. Dataset preprocessing

The dataset has numerous missing data, which may lead to error when passed as the input. To solve this problem, we eliminated all the features that had more than 90% missing values to ensure that minimum information is obtained for each feature. Then, we eliminated the records that had a high missing value. Thus, the number of patients decreased from 5,644 to 600; this number is similar to that of other studies on this dataset [15,16]. Dealing with categorical data was another problem in this dataset; because ML uses mathematical equations, all data must be passed as numeric inputs. Therefore, we used labels with a unique number for the categorical variables in every record in the columns.

3.3. Synthetic minority oversampling technique (SMOTE)

The general problem with medical data is data imbalance, which occurs when the number of patients is lower than the number of non-patients. This poses a significant challenge in ML calculations. Oversampling algorithms, in contrast to undersampling approaches, do not process majority samples but increase minority samples to boost category efficiency. The SMOTE [20] creates synthetic minority samples by interpolating a randomly chosen and homogeneous neighbouring sample, which can improve the detection rate of the minority samples. Overlapping samples, boundary samples, noise samples, and other issues may arise as a result of the rules used to generate the new samples. To overcome these problems, in [21], a combined method of cleaning and retrieval was proposed, which clears the selection boundaries before placing the sample in the minority class to eliminate the boundary sample affecting the selection. In another study, Gaussian-SMOTE was used mainly on the basis of the Gaussian function. This sample distribution is calculated by the Gaussian function to control for minority sample fitting, thus effectively solving the problem of marginal sampling.

3.4. Recursive feature elimination (RFE)

To determine the necessary classes, identifying the attributes that are appropriate or most important to the classification is essential. Generating a subset of candidates involves searching for a feature subset, which is then used as an input for the evaluation process. The feature selection algorithm begins by choosing an initial subset on the basis of three stages of subset formation: (1) The subset is initially empty, and during search, the algorithm adds individual features one by one to this subset (called forward search); (2) If the original subset is identical to the feature set of a particular dataset, inappropriate or redundant features will expire from the original section during the search (called reverse search); (3) the initial subset is randomly generated and functions are inserted or removed one by one as the search progresses [22]. A large number of features is another problem posed by this dataset. More than 100 features exist, and if all of them are used in the training stage, the processing time increases and predictability decreases. Furthermore, other problems such as duplicate features, choosing ineffective features, and high contradiction between features are encountered. To solve all these problems, we use RFE, which is one of the Sklearn methods that can work with any estimator.

3.5. Optimization of machine learning system

The working of DL and ML algorithms differs. Until recently, DL algorithms were restricted in terms of strength and complexity. With scientific revolution, larger and deeper networks have gained demand to manage large data, and machines are being used to learn, monitor, and interact with complex conditions in a better and more efficient manner than humans can. In this research, we develop and evaluate different models with laboratory data to classify COVID-19 infection. The models that we train are ANN, convolutional neural networks (CNN), and AdaBoost. The ANN is a part of DL methods and includes a series of simple, connected adaptive neural nodes to simulate the human brain. It is mainly based on complex mathematical models and advanced software tools. The CNN [17] is a class of neural networks that has hidden layers called convolutional layers and is able to detect patterns. For achieving pattern recognition, the CNN is applied in image analysis research. The model can also be used on one-dimensional sequences of data. It learns to extract features from the sequence of observations, which is in contrast to ML. AdaBoost is one of the ensemble boosting classifiers. A recurrent neural network (RNN) is a form of neural network that evaluates the flow of data through a hidden layer unit. It is determined by the results of the previous calculation series. Many areas including text processing, speech recognition, and health industry rely on sequencing data through DNA sequencing and other healthcare applications. Because several classifiers are combined to increase the accuracy, it is an iterative ensemble method. The basic idea of AdaBoost is to set the weights of classifiers and train the sample data in each iteration; in each round, greater weight is assigned to the samples that are not successfully learned in the previous iteration [23].

A common problem with DL is overfitting, which occurs when an ML algorithm picks up noise or the algorithm fits data for a long time. Batch normalisation and dropout are well-known approaches to handle these challenges; many research findings have demonstrated that they have unique strengths for improving DL. We use optimisation with the ANN and CNN because this approach can reduce the prediction time, rate of error, and overfitting and help in determining the best performing parameters.

4. Results and discussion

To overcome the obstacles of previous models in solving the data imbalance problem, we apply a new hyper model to train the DL and ML classification models. We eliminate the high-value missing data to increase the statistical power of the models. SMOTE oversampling is applied to overcome most classes in the dataset. RFE is applied to eliminate the data with less-important features and
reduce training time, while determining the most distinguishing features. In addition, the laboratory findings (features) in the dataset do not have a specific range; therefore, we replace the range of features with the minimum and maximum values, often ranging between 0 and 1. The models used in this study include the ANN, CNN, RNN and AdaBoost classifiers to diagnose whether a person has COVID-19 or not. In addition, grid search optimises the hyperparameters in our models. Table 1 presents the hyperparameters of the models.

We examined the performance of all models by splitting the data into training and testing sets in the ratio 80:20. Based on experiments, we built the system architectures for the ANN, CNN, and RNN models. Further, we developed and optimised via the grid-search optimisation approach to assign hyperparameters to all DL models. Epoch is a hyperparameter that controls the number of times a DL algorithm runs over the full training dataset, i.e., batch size. Because one epoch is too large to feed the network at once, which will produce a high loss in the learning, we split the training dataset into several smaller batches. Further, dropout was used as another DL hyperparameter to minimise overfitting by dropping some neurons randomly from the hidden layers. Because the dataset has two classes (0,1), binary cross-entropy was used to calculate the loss in each epoch.

As shown in Table 2, the metrics F1-score, accuracy, and recall of all models were >91.00%. Accuracy determines the closeness of sample parameters to the characteristics of a population [22] and is obtained as the number of correct classifications divided by the total number of correct cases. The total number of acceptable ratings is divided by the total number of instances to calculate the model output. The best accuracy of 98% was obtained using the ANN. Precision indicates all the samples that have been classified as positive; a precision of 1 is considered ideal and indicates the most accurate and precise classification. Recall indicates the proportion of the true positive samples among all positive samples. Similar to accuracy, it should ideally be 1. The F1 score is a single statistic that combines the precision and recall metrics and indicates the effectiveness of the operation of a model with imbalanced datasets.

Table 2 indicates the overall performance of the dataset obtained after preprocessing. In addition, five metrics were used for evaluation: accuracy, precision, F1 score, recall and AUC score. The accuracy of all DL and ML models reached a minimum of 91.00%. The ANN model showed the best performance, with accuracy reaching 98% and highly balanced data with an F1 score reaching 0.98. In addition, the training and testing accuracy graphs are obtained with the training loss and testing loss graphs because they effectively demonstrate whether the models encounter problems of underfitting or overfitting. As shown in Figs. 3 and 4, both graphs indicate that all models have a good fitting through a number of epochs. The CNN was noted as the second-best model with 96% accuracy, and the RNN was the third best model. The reason for the excellent performance of the ANN is that it is a very robust method against noise in training data and can implicitly reveal complex nonlinear relationships between dependent and independent variables [24].

The AUC is most often used in binary classification problems to determine the best model for class predictions. The number of true positives is divided by the number of false positives to obtain the AUC score [17]. The higher the score, the better is the model in predicting the classes. Therefore, a value between 80% and 90% is considered sufficient, although a value > 90% is ideal. The ANN model achieved an ideal score of 99%. Moreover, other models’ scores were excellent, with all being > 91%. According to the AUC results, COVID-19 prediction could be achieved using both DL and ML methods. As a result, the AUC score plays an important role in medical datasets because it provides a clear explanation to distinguish infected people from healthy people. Fig. 2 shows the analysis graph of AUC scores for both DL and ML models.

As shown in Fig. 3, visualising training loss (orange line) versus test loss (blue line) over the number of epochs is an effective means to determine whether the model is trained adequately or is undertrained or overtrained. This study demonstrates that the ANN model is the best for minimising loss and obtaining reasonable results with the new data and not only with data that have already been trained. For effective training of an ML model, overfitting must be avoided, which occurs when the learning algorithm fits the training data well, leading to high training accuracy but poor test accuracy. Therefore, the test accuracy should be close to the training accuracy; to reduce the difference between the two, we use the dropout hyperparameter. In Fig. 4, blue and orange lines represent the accuracy of the training data and testing data, respectively. As shown in the figure, all DL models fit well between the testing and training datasets, with the accuracy of training and testing being relatively close.

Table 3 shows the results of the compilation of the proposed method and other latest methods for comparison. In studies [15] and [16] using ML, SVM and MLP classifiers, respectively, were used for classification. In [14], along with ML, four algorithms with SHAP were used to select the most accurate features for predicting COVID-19 patients. In addition, in [17], DL combined with the CNN-LSTM was used. By contrast, in this study, we used DL and ML models and developed a new system that can handle unbalanced datasets such as the COVID-19 dataset. Our method provided better prediction accuracy than the existing methods.

| Parameters               | ANN     | CNN     | RNN     |
|--------------------------|---------|---------|---------|
| Number of neurons        | 64,32,16,8 | 64,32,16,8 | 256,128,64,32 |
| Number of layers         | 1,2,3,4 | 1,2,3,4 | 1,2,3,4 |
| Activation function      | Relu    | Relu    | Relu    |
| Loss function            | “Binary Cross-entropy” | “Binary Cross-entropy” | “Binary Cross-entropy” |
| Number of Epochs         | 100     | 100     | 20      |
| Optimizer                | Adam    | Adam    | Adam    |
| Dropout                  | 0.15    | 0.20    | 0.15    |
| Batch size               | 10      | 16      | 65      |
Table 2
Performance results for all models.

| Classifier   | Accuracy | Precision | F1-Score | Recall | AUC  |
|--------------|----------|-----------|----------|--------|------|
| ANN          | 0.98%    | 0.97%     | 0.98%    | 0.99%  | 0.99%|
| CNN          | 0.96%    | 0.95%     | 0.96%    | 0.98%  | 0.99%|
| RNN          | 0.94%    | 0.94%     | 0.91%    | 0.97%  | 0.97%|
| Ada Boost    | 0.91%    | 0.91%     | 0.91%    | 0.91%  | 0.91%|

Fig. 2. AUC graph analysis for all models.
5. Limitations

The major limitations of this study are the small number of patients in the dataset and the lack of some laboratory results. However, despite the insufficiency of laboratory results, the prediction accuracy of 91%–98% could be achieved. In addition, the information was imbalanced, which was balanced using the SMOTE. The performance of the models can be improved further by using a larger dataset. In addition, other combinations of feature selection and unbalanced learning algorithms should be investigated within the framework described in this study to enhance the accuracy of prediction of COVID-19 patients.

6. Conclusion and Future Work

Early detection of COVID-19 is crucial for treating patients in a timely manner and controlling the pandemic. Laboratory tests to screen patients have proven useful because clinical tests are relatively cheaper, more expensive, and readily available in most treatment centres. In this study, we first provided an overview of the existing methods for detecting COVID-19 based on routine laboratory and clinical data to inspire researchers to develop effective predictive models to combat the deadly disease. We used COVID-19 data as an example of an unbalanced dataset for predicting the outbreak of the disease based on laboratory results. To achieve balance in the dataset, the SMOTE was used, which increases the minor class by generating new data. Further, feature selection
**Table 3**
Comparison between evaluation results of different models.

| Study     | Classifier               | Accuracy | F1-Score | AUC  | Year of Publication |
|-----------|--------------------------|----------|----------|------|---------------------|
| This Work | ANN                      | 98%      | 98%      | 99%  | -                   |
| [14]      | Machine Learning         | 91%      | 85%      | 92%  | 2022                |
| [15]      | MLP                      | 93%      | -        | -    | 2020                |
| [17]      | CNNLSTM, ANN             | 92%      | 93%      | 90%  | 2020                |
| [16]      | SVM                      | 80%      | -        | -    | 2020                |

Fig. 4. Difference between training and test accuracy.
was used to remove ineffective features and reduce the training time. The new laboratory data that were selected by RFE were analysed using the DL and ML models. The models were validated by splitting the dataset into training and testing sets and optimising the parameters to reduce the computation time and enhance the classification accuracy. The clinical prediction performance of most of the classification algorithms was good, with the ANN showing the best evaluation metrics of 98% accuracy. The CNN optimisation helped in achieving 96% accuracy and 99% AUC for both ANN and CNN. In addition, the AdaBoost provided 91% accuracy and 91% AUC. Furthermore, by comparing the results with those of other existing methods, the efficacy and reliability of classifiers in diagnosis were determined. Early diagnosis of COVID-19 can be introduced in future research by using a prediction model based on the stacked generalisation model and data augmentation.

Declaration of Competing Interest

The authors whose names are listed immediately below certify that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers’ bureaus; membership, Employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or Materials discussed in this manuscript.

References

[1] Harapan H, Itoh N, Yufika A, Wisaward A, Koon S, Te H, Megawati D, Hayati Z, Wagner AL, Mudatsir M. Coronavirus disease 2019 (COVID-19): a literature review. J Infect Public Health May 2020;13(5):667–73. https://doi.org/10.1016/J.JIPH.2020.03.019.
[2] Jiehao C, Jin X, Daojiong L, Zhi Y, Lei X, Zhenghui Q, Yuelhua Z, Hua Z, Ran J, Pengcheng L, Xiangshi Y, Yanling G, Aimei X, He T, Hailing C, Chuning W, Jingling L, Jianhe W, Mei Z. A case series of children with 2019 novel coronavirus infection: clinical and epidemiological features. Clin Infect Dis Sep 2020;71(6):1547–51. https://doi.org/10.1093/cid/ciaa198.
[3] Manigandan S, Wu MT, Ponnamasy VK, Raghavendra VB, Pugazhendhi A, Brindhadevi K. A systematic review on recent trends in transmission, diagnosis, pathogenesis and imaging features of COVID-19. Process Biochem 2020;88(May):233–40. https://doi.org/10.1016/j.procbio.2020.08.016.
[4] Naseem M, Akhund R, Arshad H, Ibrahim MT. Exploring the potential of artificial intelligence and machine learning to combat COVID-19 and existing opportunities for LMIC: a scoping review. J Prim Care Community Heal Sep. 2020;11. https://doi.org/10.1177/2150132720963634.
[5] Taryani MH. Applications of artificial intelligence in battling against Covid-19: a literature review. Chaos, Solitons Fractals Oct. 2020;142. https://doi.org/10.1016/J.CHAOS.2020.110338.
[6] Poon SC, McNally MC. Machine learning. Pattern Recognit Mach Learn Dec. 2006. https://doi.org/10.1097/978-0-387-45528-0.
[7] Shinde GR, Kalamkar AB, Mahalle PN, Dey N, Chaki J, Aboul, Hassanien E. Forecasting models for coronavirus disease (COVID-19): a survey of the state-of-the-art. SN Comput Sci Jun. 2020;4(1):1–15. https://doi.org/S2497-020-00209-9.
[8] Elgamael ZM, Yasini NM, Sabri AQM, Sliwiarw R, Tushidt M, Jarrahi H. Improved equilibrium optimization algorithm using elite opposition-based learning and new local search strategy for feature selection in medical datasets. Computur J. 2021;9(6):68. https://doi.org/10.3390/COMPUTATION9060068, 2021, Vol. 9, Page 68.
[9] Abdolrahimzadeh Fard H, Mahmudi-Azer S, Abdolzahraa Yaqoop Q, Sahetian G, Imanpour P, Shayyan Z, Bolandparvaz S, Abbasi HR, Aminnia S, Salimi M, Mahmoudi MM, Paydar S, Borazjani R, Taheri Akerdii A, Zare M, Shayyan L, Sasanii M. Comparison of chest CT scan findings between COVID-19 and pulmonary contusion in trauma patients based on RSNA criteria: Established novel criteria for trauma victims. Chinese J Traumatol Jan. 2022. https://doi.org/10.1016/J.CJTEE.2022.01.004.
[10] Faris H, Abukhurma R, Almanaseer W, Saadeh M, Mora AM, Castillo PA, Aljahri J. Improving financial bankruptcy prediction in a highly imbalanced class distribution using oversampling and ensemble learning: a case from the Spanish market. Prog Artif Intell 2020;9(1):31–53. https://doi.org/10.1016/j.bmit.2020.100374.
[11] Liu J, Zhang X, Tong Q, Li W, Wang B, Sutter K, Trilling M, Lu M, Dittmer U, Yang D. Overlapping and discrete aspects of the pathology and pathogenesis of the emerging human pathogenic coronaviruses SARS-CoV, MERS-CoV, and 2019-nCoV. J Med Virol May 2020;92(5):491–4. https://doi.org/10.1002/JMV.25709.
[12] X. Deng, H. Shao, L. Shi, X. W.-C. M. in, and undefined 2020, Pattern recognition and machine learning. Pattern Recognit Mach Learn Dec. 2006. https://doi.org/10.1007/978-0-387-45528-0.
[13] Mahmoudi MM, Paydar S, Borazjani R, Taheri Akerdii A, Zare M, Shayyan L, Sasanii M. Comparison of chest CT scan findings between COVID-19 and pulmonary contusion in trauma patients based on RSNA criteria: Established novel criteria for trauma victims. Chinese J Traumatol Jan. 2022. https://doi.org/10.1016/J.CJTEE.2022.01.004.
[14] Chadaga K, Chakraborty C, Prabhu Srikantan, Umakanth S, Bhat V, Sampathila N. Clinical and laboratory approach to diagnose COVID-19 using machine learning. Interdiscip Sci Comput Life Sci Feb. 2022;1:1–19. https://doi.org/10.1119/ACCESS.2022.3142522.
[15] Mondal MRH, Bharati S, Podder P, Podder P. Data analytics for novel coronavirus disease. Informatics Med Unlocked Jan. 2020;20:100374. https://doi.org/10.1016/j.imu.2020.100374.
[16] Jiang X, Coffee M, Bari A, Wang J, Jiang X, Huang J, Shi J, Dai J, Cai J, Zhang T, Wu Z, He G, Huang Y. Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. Computur J. 2020;53(s1):537–51. https://doi.org/10.1093/computermagazine/202001691.
[17] Alakus TB, Turkoglu I. Comparison of deep learning approaches to predict COVID-19 infection. Chaos, Solitons Fractals Nov. 2020;140:110120. https://doi.org/10.1016/J.CHAOS.2020.110120.
[18] Samuel J, Ali GGMN, Rahman MM, Eswai E, Samuel Y. COVID-19 public sentiment insights and machine learning for tweets classification. Information Jun. 2020;11(6). https://doi.org/10.3390/information11060314.
[19] Khan MA, Abbas S, Khan KM, Ghamdi MAA, Rehman A. Intelligent forecasting model of covid-19 novel coronavirus outbreak empowered with deep extreme learning machine. Computur J. 2020;63(3):1329-42. https://doi.org/10.1093/computermagazine/202001155.
[20] Geetha R, Sivasubramanian S, Kaliappan M, Vimal S, Annamalai S. Cervical cancer identification with synthetic minority oversampling technique and PCA analysis using random forest classifier. J Med J. 2019;43(9):1–19. https://doi.org/10.1016/J.MJMS.2019.012-14026.
[21] Raghuvanshi BS, Shukla S. SMOTE based class-specific extreme learning machine for imbalanced learning. Knowledge-Based Syst Jan. 2020;187:104814. https://doi.org/10.1016/J.KNOSYS.2019.06.022.
[22] Qasim HM, Ata O, Azam Ansari M, Alomary MN, Alghamdi S, Almehmadi M, Atrusi A, Guerra A, Schirinzii T. Hybrid feature selection framework for the parkinson imbalanced dataset prediction problem. Medicina Nov. 2021;57(11):1217. https://doi.org/10.3396/MEDICINA57111217. 2021, Vol. 57, Page 1217.
[23] Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, Mishra R, Pillai S, Jo O. COVID-19 patient health prediction using boosted random forest algorithm. Front Public Heal (August)2020;8:1–9. https://doi.org/10.3389/fpubh.2020.000357.
[24] Hajiabadi H, Babaiyan V, Zahidhadeh D, Hajishad M. Combination of loss functions for robust breast cancer prediction. Comput Electr Eng Jun. 2020;84: 106624. https://doi.org/10.1016/J.COMPELECENG.2020.106624.
Roa’a Mohammedqasem is a Ph.D. student at the Institute of Electrical and Computer Engineering (ECE), at the University of Altinbas in Turkey. She received her Bachelors in 2007 and her Masters in 2018. Her interests include machine learning, artificial intelligence, data mining, and IoT.

Hayder Mohammedqasim is a Ph.D. student in Data since at Altinbas University in Turkey. He is interested in machine learning and its applications in important situations such as medical diagnosis, with a focus on overcoming the many problems in machine learning and the intrinsic limitations of a neural network, particularly its comprehension.

Oguz Ata received the B.Sc. degree in computer engineering from Sakarya University, in 2004, the M.Sc. degree in computer engineering in 2008, and the Ph.D. degree in computer engineering, in 2012. He works at Altinbas University as an assistant professor doctor. His research interests include wireless sensor network, data mining, artificial intelligence, image processing, and machine learning.