Noname manuscript No.
(will be inserted by the editor)

Analysing And Predicting COVID-19 AI Tracking Using Artificial Intelligence

IYAS ALODAT
Jerash University
Computer Science and Information Technology
E-mail: eyas.odat@jpu.edu.jo

Received: date / Accepted: date

Abstract In this paper, we will discuss prediction methods to restrict the spread of the disease by tracking contact individuals via mobile application to individuals infected with the COVID-19 virus. We will track individuals using bluetooth technology, then we will save information in the central database when they are in touch. Monitoring cases and avoiding the infected person helps with social distance. We also propose that sensors used by people to obtain blood oxygen saturation level and their body temperature will be used beside bluetooth monitoring. The estimation of the frequency of the disease is based on the data entered by the patient and also on the data gathered from the users who entered the program on the state of the disease. In this study, we will also propose the best way to restrict the spread of COVID-19 by using methods of artificial intelligence to predict the disease in Jordan using Tensorflow.

Keywords Deep Leaning · Mobile Application · SPO2 Sensors · Tensorflow

1 Introduction

Via tiny drops in the nose or mouth, people may catch infection with the COVID-19 epidemic from whoever has the virus mainly from person to person, which are extracted when a person with COVID-19 coughs, sneezes or talks. Such drops are surprisingly heavy, do not travel far and fall quickly to the ground. People are catching COVID-19 if they inhale these droplets from a person with the virus. That is why it is better to keep at least 1 meter away from everyone. These droplets can land on various surfaces around a person

F. Author
Jerash University
Computer Science and Information Technology
E-mail: eyas.odat@jpu.edu.jo

such as tables, door knobs, and surfaces. Individuals can become infected by touching these items or surfaces, and then touching their eyes, nose , or mouth. This is why it is important to wash your hands regularly and to clean them with an alcohol-based antiseptic. Subsequently, all these preventive measures are exhausting among infected people and other new approaches are needed to limit the spread of COVID-19, which can be accomplished using modern technology.

COVID-19 spreads mainly through a person who coughs or shows symptoms such as fever or fatigue, expelled by respiratory droplets. There are only minor symptoms in many people with COVID-19, and this is particularly true in the early stages of the disease. Covid-19 disease may be captured by a person who suffers from mild cough and does not feel anxious about it. Some reports have indicated that the virus can be transmitted by people with no symptoms.

At the beginning of 2020, coronavirus disease (COVID-19) has spread around the world. It is extremely infectious and normally provoke breathing difficulties or kidney failure in serious situations [1], [2], [3], [4]. The World Health Organization (WHO) declared the epidemic of COVID-19 as a "Public Health Emergency of International Significance" (PHEIC) on 30 January 2020.

Reverse Transcription of the Polymerase Chain Reaction (RT-PCR) has consistently documented the condition. However, it has been noted that the sensitivity of RT-PCR may not be sufficiently high for the early detection and treatment of presumptive patients [5], [6]. Non-invasive imaging approach such as computerized tomography (CT) scan can detect some of the characteristics associated with lung COVID-19 [7], [8]. Recently, the use of deep learning technology by artificial intelligence (AI) has shown great success in many areas in improving and predicting the best results, for example in the field of medical imaging due to a high ability to extract [9], [10], [11]. Deep learning has also been used to diagnose and distinguish between bacterial and viral pneumonia in childrens chest X-rays [12], [13]. Attempts have also been made to discover the different imaging properties of chest CT imaging [14],[15].

The home quarantine for a duration of 14 days is one of the methods used to diagnose the disease and restrict the spread of this outbreak, Due to this fact, this matter is important at the beginning of the discovery of the disease, any person who has been exposed to an infected person must quarantine himself for 14 days. In order to complete this process and to better identify the exposure process to people who are infected, we have been working on a special software that works on mobile. we can identify the infected people via communication through Bluetooth technology; when people have a direct interaction with a patient(less than 1 meter distance) this interaction will be monitored. Here, individuals need to trust the program itself and regularly fill out the current questionnaire to recognize the condition of the user and the rate of his infection to ensure better outcomes in the future.

2 Methodology

Since it is well known that data mining techniques (e.g. artificial neural networks) for deep learning models (including classification and clustering tasks) are capable of discovering patterns and groups of potential data, such as classifying an infection of COVID-19. This study highlights a model for applying data mining to analyze and predict a COVID-19 tracking service in mobile technology.

Based on that, it is required to link the analysis of the tracking service with it's prediction capability. Also, the design components of a preferable development of a mobile tracking application. Hence, using data mining views a multidimensional perspective. This might lead to a better decision such as, what is the best configuration of designing a successful mobile tracking application, what is the critical component that impact the efficiency of the mobile tracking application, or even the component that impact an effective interaction with the application. In general, data mining techniques are to discover patterns/factors that have the potential to become a fully integrated mobile tracking application. For more clarification of the capabilities and applications of data mining techniques, please refer to related studies of deep learning [21][22][23].

We are using bluetooth technology in mobile application; also we are using sensors that are connected with mobile application.

All data will be collected and send to central database for analysing and processing to have results that are suitable with artificial intelligence methods.

2.1 Data Collection

Through using mobile application with bluetooth technology, we will collect all data about the people which are contacted with the infected person, and through the application we use a questionnaire to capture information on the patient's condition. Also, when the patient measures the body temperature and the oxygen saturation level in the blood periodically, then in this case we will predict the pathological state of the epidemic at the state level.

Through the questionnaire in the mobile application, the person will answer the questions. As shown in figures 1 and 2, we use the questionnaire to predict epidemiological conditions, while sensors were used to obtain personal information.

These data will be used for the method of prediction by means of artificial intelligence, as we use the questionnaire in the software to determine the nature of the disease in the patient by asking about the symptoms of the disease. In addition, we use the questionnaire to determine whether the person is suffering from a chronic disease, or to determine whether the person comes from a country with a high risk of an epidemic; according to the classification adopted

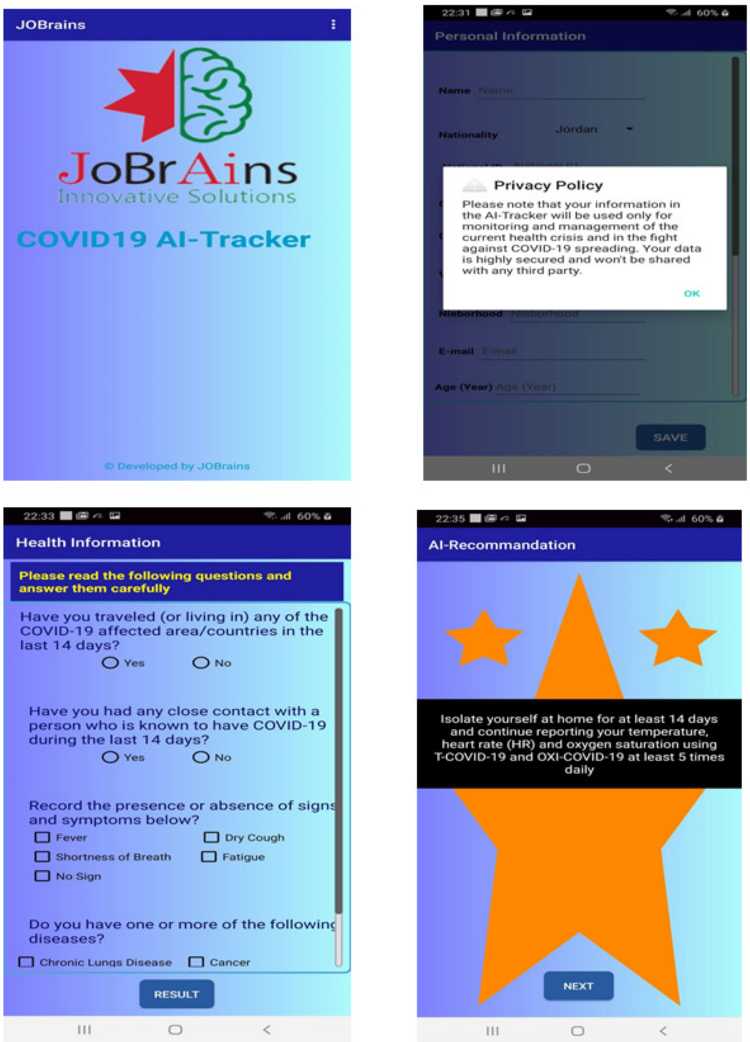

Fig. 1 The Application of COVID-19 AI Tracker: Personal Information and Questionnaire

globally. Also, the sensors were used to measure the person's body temperature and to measure the person's blood oxygen level.

2.2 Experimental Analysis

After the data collection phase, a set of methods have been selected to predict the disease status in the country, using Tensorflow to display the results of algorithms used to predict the epidemiological situation in the country, and then the phase of selecting the best algorithm for artificial intelligence algo-

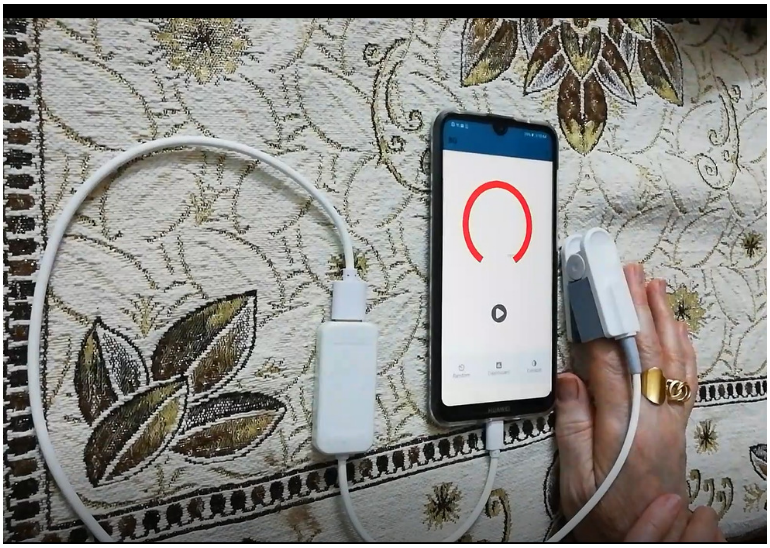

Fig. 2 Using Sensor for Measuring Blood Oxygen Level And Temperature

rithms to predict epidemiological status was accomplished. In this article, we will compare the best approaches we can use from artificial intelligence to determine the intensity of the epidemic.

3 Background of Deep Learning Models

In this section, we will review the methods used for artificial intelligence operations to predict the future state of the processed data set.

3.1 Supervised Models

Learning data provides an inferred function that can be used to map later training instances. An optimized scenario will allow the algorithm to immediately recognize training data for various visual situations. This helps the learning algorithm to generalize, in a specific way, from the training data. The right induction can be achieved by the following steps: Deciding the sort of examples of training, such as the recognition of handwriting in a letter or phrase typed, taking the set of training representing the actual use of the function in a letter or sentence as a sample, and then determination of the input feature of the learned function containing a number of object descriptive characteristics.

Due to the dimensionality, the number of features may not be too high, but this will contain enough information to reliably predict performance, determine the structure of the work obtained and the corresponding learning algorithm, and analyze the accuracy of the work learned. The performance of the work should be measured in a test group separate from the training session, after parameter adjustment and learning.

3.1.1 Classic Neural Networks

Multilayer Perception (MLP) is a basic artificial neural network. It consists of much more than one predictor. It takes an input for receiving a message, an output layer that makes an input a conclusion or an estimation, and a random number of hidden layers between the two layers that represent the true MLP computational engine. Any continuous work can be calculated by a single hidden layer of MLPs. In three supervised learning disabilities, layered cognition is often applied;they train a set of input-output pairs and learn how to model the relation (or dependence) between those inputs and outputs. In order to eliminate errors, training involves the adjustment of model parameters, weights and biases. Back-propagation is used to correct weight and bias relative to the error. MLP can measure the error directly in a number of ways, including the Root Mean Square Error (RMSE) [18]. Station networks are close to tennis courts or table tennis courts. Like MLPs, they are basically engaged in two actions; a steady forward circular motion. You can think of ping pong from predictions and answers as a kind of enhanced science, since every prediction is a test of what we believe we understand, and then every reply is a comment that allows us to know how wrong we are.

3.1.2 Convolutional Neural Networks (ConvNet / CNN)

ConvNet is a deep learning algorithm that also can take an input data, assign meaning (trainable weight values) to various aspects / objects in the image, and can discriminate them from each other. ConvNet requires less data preprocessing than other classification methods. Whereas the moderators are designed by hand in primitive methods, ConvNets have the ability to learn these moderators / attributes with adequate experience. The ConvNet model is very similar to that of the neuron network of interactions in the human mind and is influenced by the peripheral stimulation control. Individual neurons react to changes only in a small area of the visible spectrum known as the receptive field. A collection of these fields intersects to represent the full visible area [19].

3.1.3 Recurrent Neural Networks (RNNs)

RNN models are mostly used in the area of natural speech language processing detection. A repetitive neural network knows the history, and it's decisions are shaped by what it has learned from the history. Note: Simple forwarding networks often "remember" information, but they know information they learned during training. For example, an object trained model learns what "1" looks like during training and then utilizes that analysis to inform objects in output. Although RNNs learn by looking while training, they often understand details

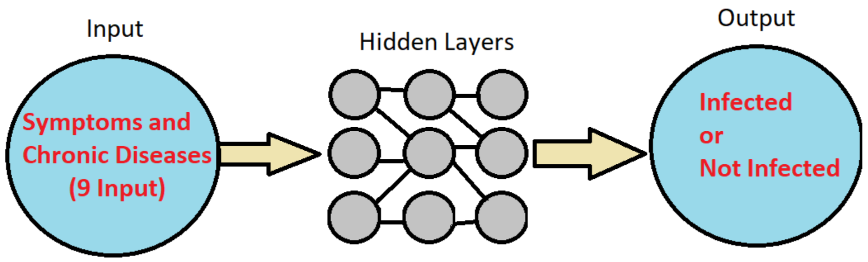

Fig. 3 Artificial Neural Networks Architecture For COVID-19

that were learned from previous inputs when generating outputs. RNNs may take one or more input data and generate one or more output layer, and output is determined not only by weights applied to inputs such as normal NN, but also by the "hidden" state vector representing the context-based ex-output (input). As a consequence, the same input will generate different outputs depending on the previous inputs in the series [20].

• As shown in figure 3, the neural network consists of 9 inputs taken from the application which are symptoms and chronic disease and personal information data, we used 3 hidden layers to predict if the person is infected or not.

• The methods we used to conclude the best way to predict the general state of the epidemic include: 1.support-vector networks 2.Kernel SVM 3.K-Nearest Neighbors 4.Logistic Regression 5.Decision Tree Classification 6.Naive Bayes methods 7.Random forest algorithm 8.Neural Network TensorFlow.

TensorFlow incorporates a variety of machine learning and deep learning (neural networking) models and makes them useful by means of a suitable metaphor. Python is used to provide a convenient front-end API for designing device applications while running those applications in high-performance C++ applications.

TensorFlow 2.0 has redesigned the code in a number of ways, based on user feedback, to make it simpler to manage (e.g. using the Keras API for model training) and more effective. Distributed testing is easier to run compared to the new API, and TensorFlow Lite support makes it possible to deploy models on a wide variety of platforms. We will apply the data collected through the mobile application to machine learning using TensorFlow, in order to deduce the future epidemiological situation.

4 Results and discussions

Data center as shown in table 1, the first column represents the first person, and the second column represents the group of people in contact with the first

Table 1 Data collected from mobile application which is sent to central database.

Bluetooth1	Bluetooth2	Q1	Q2	Q3	Q4	O2	Temp.
X	Y	1	0	1	0	97	36
X	Z	1	0	1	0	97	36

person, the remaining columns represent answers from the questionnaire by the mobile application, as well as the percentage of oxygen in the blood taken via the mobile phone and the body temperature of the person.

The data collected via the mobile application via the central database shown in table 1 was intended to resolve the epidemic in order to predict the current and potential epidemiological situation in general, and to recognize the contacts of the infected individual, as well as the infected people themselves.

The criteria that we have relied on in order to choose the best algorithm through the Tensorflow are:

The mean squared error, which is the difference between the expected values and the actual value, where in the implemented algorithms it will be obvious to us which values are the best. An indicator of the precision of the test is the F1 value.

F1 value is determined by the test accuracy criterion and the recall criterion, where the accuracy is the number of positive results correctly identified, divided by the number of all positive results, including those not correctly identified. The recall is the number of positive results correctly identified, divided by the number of all tests that should have been identified as positive. So here, the harmonic average of precision and recall is F1.

The Roc Auc Score is the ROC curve area (AUC) as shown in figure4, a numerical value that gives you an impression of the algorithm 's output applied to the data we have. ROC curves are commonly used to graphically display the connection / trade-off of each possible test interval or group of tests between clinical and specific sensitivity. Furthermore, the area under the ROC curve provides an idea of the usefulness of using the tests in the questionnaire.

Finally, the standard deviation is the most commonly used value of statistical dispersion measurements to quantify the degree of statistical dispersion, i.e. to show the extent to which the value fields within the statistical data set expand. Which is done within a set of values to give the best outcome possible.

The table 2 shows the effects of the algorithms applied to the data collected. The standard deviation tends to be approximately equal to all algorithms in the last column, meaning that the dispersion criterion for these algorithms is very similar, so we are not going to follow this criterion in the selection process.

In the first column (mean squared error), we see that each of the algorithms took convergent values with the exception of the (Kernel SVM) algorithm, which has a lower value than the rest of the algorithms, so we will use this

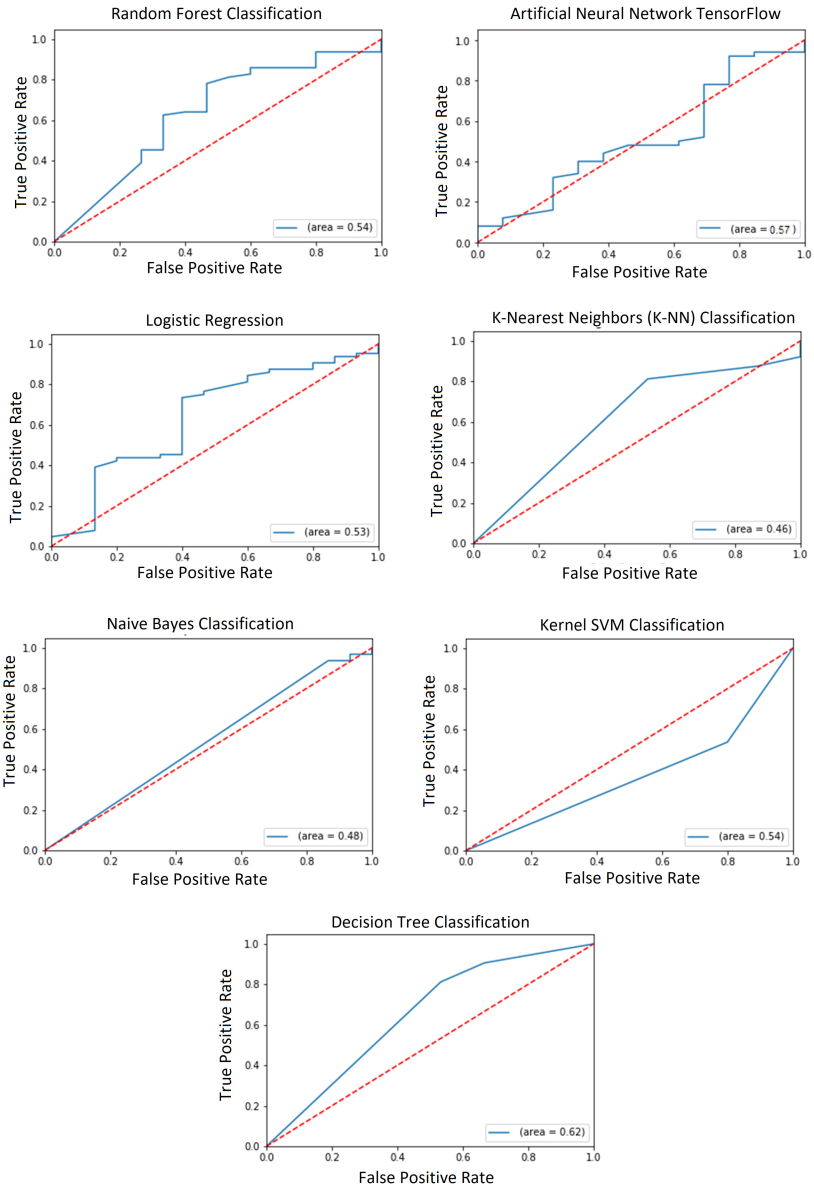

Fig. 4 ROC curve results for all algorithms

Table 2 Parameters of Tensorflow Deep Learning Models

	Mean Squared Error	F1 Score	Roc Auc Score	Standard Deviation	precision
(SVM) model	0.2152	0.8759	0.5354	0.0582	0.8219
Kernel SVM	0.1266	0.9315	0.5428	0.023	0.8831
K-Nearest Neighbors	0.2532	0.8551	0.4609	0.0437	0.7973
Logistic Regression	0.2278	0.8676	0.5276	0.0596	0.8194
Decision Tree	0.2025	0.8788	0.6198	0.0334	0.8529
Naive Bayes	0.2152	0.8794	0.4844	0.0211	0.8052
Random Forest	0.2532	0.8485	0.5375	0.061	0.8235
NN TensorFlow	0.2381	0.8571	0.5654	0.0505	0.8182

value into account in our analysis, considering the importance of this criterion in showing the best quality of the prediction process, the remaining two criteria also have the same value.

We noticed that when evaluating all the values in the table, the two algorithms (Kernel SVM) and (Decision Tree Classification) are the ones that we will apply the comparison method between them as the best predictive result that can be used in the future. But after looking at figure 4, we can see that the SVM kernel in the ROC curve defines which means would be useful algorithms that can be used to predict performance.

5 Conclusion

After the data collection from mobile applications which includes: Contact information through Bluetooth, answers taken from the questionnaire and the personal information data collected from sensors. We conclude that the final outcome of our study shows that it is better to use the algorithm (Kernel SVM) because it will give us the best result to predict the future values of COVID-19 based on the mentioned criteria (MSE, F1).

As for the roc_auc_score criterion, the decision tree classification is better than others, but this does not give it the preference for predicting the future, especially that the algorithm (Kernel SVM) has outperformed other algorithms in this study in the two mentioned criteria and is close to the value in relation to the roc_auc_score criterion. So here, based on the existing method, we recommend using the (Kernel SVM) algorithm because it will give the best results for predicting the future values of the COVID-19 and will help limit the spread of this epidemic in a better way.

References

1. Chen N, Zhou M, Dong X, et al. . Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. doi: 10.1016/S0140-6736(20)30211-7. The Lancet 2020 Jan 30.
2. Wang D, Hu B, Hu C, et al. . Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. Jama 2020.
3. Li Q, Guan X, Wu P, et al. . Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. The New England Journal of Medicine 2020.
4. Holshue ML, DeBolt C, Lindquist S, et al. . First Case of 2019 Novel Coronavirus in the United States. doi:10.1056/NEJMoa2001191.New England Journal of Medicine 2020 Jan 31.
5. Ai T, Yang Z, Hou H, et al. . Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. 10.1148/radiol.2020200642. 2020.
6. Fang Y, Zhang H, Xie J, et al. . Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. 10.1148/radiol.2020200432. 2020
7. Chung M, Bernheim A, Mei X, et al. . CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). : 200230. 2020
8. Huang C, Wang Y, Li X, et al. . Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. doi: 10.1016/S0140-6736(20)30183-5. The Lancet 2020 Jan 24
9. Xia C, Li X, Wang X, et al. . A Multi-modality Network for Cardiomyopathy Death Risk Prediction with CMR Images and Clinical Information. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019.
10. Kong B, Wang X, Bai J, et al. . Learning tree-structured representation for 3D coronary artery segmentation. 80: 101688, Computerized Medical Imaging and Graphics 2020
11. Ye H, Gao F, Yin Y, et al. . Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. European Radiology, 29: 6191–6201, 2019.
12. Kermany DS, Goldbaum M, Cai W, et al. . Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. 172, 1122–1131,Cell 2018.
13. Rajaraman S, Candemir S, Kim I, et al. . Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. 1715, Applied Sciences 8.10 (2018).
14. Depeursinge A, Chin AS, Leung AN, et al. . Automated classification of usual interstitial pneumonia using regional volumetric texture analysis in high-resolution CT. Investigative Radiology 50.4 : 261, (2015).
15. Anthimopoulos M, Christodoulidis S, Ebner L, et al. . Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Transactions on Medical Imaging 35.5 : 1207-1216,(2016).
16. Stuart J. Russell, Peter Norvig (2010) Artificial Intelligence: A Modern Approach, Third Edition, Prentice Hall ISBN 9780136042594.
17. Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, Foundations of Machine Learning, The MIT Press ISBN 9780262018258,(2012).
18. Bienvenido-Huertas, D., Rubio-Bellido, C., Pérez-Ordóñez, J. L., & Moyano, J. Optimizing the evaluation of thermal transmittance with the thermometric method using multilayer perceptrons. Energy and Buildings, 198, 395-411,(2019).
19. Tan, M., & Le, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946,(2019).
20. Liu, C., Li, Z., Hu, W., Xing, L., & Peng, H. Range and dose verification in proton therapy using proton-induced positron emitters and recurrent neural networks (RNNs). Physics in Medicine & Biology, 64(17), 175009,(2019).
21. Hossain, M. S., & Muhammad, G. (2019). Emotion recognition using deep learning approach from audio–visual emotional big data. Information Fusion, 49, 69-78.
22. Ardabili, S., Mosavi, A., Dehghani, M., & Várkonyi-Kóczy, A. R. (2019). Deep learning and machine learning in hydrological processes climate change and earth systems a systematic review. In International Conference on Global Research and Education, 52-62, Springer, Cham.

23. Banerjee, A. K., & Arora, N. (2020). Machine learning techniques in biological data classification and clustering: Initiation of a scientific voyage. Journal of PeerScientist, 2(1), e1000011.