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.
8.1 Introduction

The coronavirus pandemic, also known as COVID-19 pandemic (Wikipedia, 2020) created a world-wide chaos since its advent in December 2019 in Wuhan, China. With over 1,06,94,288 confirmed cases across the world, the virus is spreading at an exponential rate. The World Health Organization stated that COVID-19 is by far the most severe global health emergency and declared it as a Public Health Emergency of International Concern on January 30, 2020 (Wikipedia, 2020). As of August 6, 2020, more than 188 countries and territories have reported more than 18.7 million cases of COVID-19. The transmission rate of the virus is significantly high and spreads mainly via respiratory droplets from coughing, sneezing, talking, and also close contacts. Like a typical airborne disease, the droplets remain suspended in the enclosed spaces, which also causes the transmission. The virus can even spread from a person who does not have any symptoms, which makes COVID-19 such a dangerous entity.

From the overall cases so far, the most common symptoms include fever, cough, and shortness of breath (Webmd, 2020). The conditions of the patient deteriorate in cases of pneumonia and acute respiratory distress syndrome (United States Centers for Disease Control and Prevention (CDC), 2020a). The onset of symptoms takes typically around five days from the time of exposure but may range from 2 to 14 days (United States Centers for Disease Control and Prevention (CDC), 2020b; Velavan and Meyer, 2020). Many medical organizations over the world are testing their vaccines to
overcome this virus. So far, the standard method of diagnosis includes the real-time reverse transcription polymerase chain reaction, that is., rRT-PCR (Zu et al., 2020), which is performed on throat-swab specimens. To restrict the contamination of the coronavirus certain preventive measures are recommended, which include proper sanitization, wearing protective face masks in public, ensuring proper distance from other people, and maintaining self-isolation for people who suspect any hint of the coronavirus symptoms.

Till now, there is an absence of standard methods for predicting the risk level of a patient based on their early and mild symptoms. Since its advent, the virus has changed its character and evolved quite often, which has increased the unpredictability to judge the infection taking only a handful of symptoms. Therefore the authors in this chapter have attempted to predict COVID-19 tasking a list of symptoms and conditions based on their severity, frequency, and impact on the patients. The risk assessment has been categorized into three different classes, which are Not Infected, Mildly Infected, and Severely Infected. Keeping in mind the evolving nature of the virus, we have applied the Artificial Neural Network (ANN) to classify the risk into the mentioned classes. The use of ANN reduces the judgment of error while predicting the outcome, which again is an important feature taking into consideration the severe effect of coronavirus.

The chapter is constructed as follows: section 8.2 deals with related studies followed by the fundamental symptoms and conditions responsible for COVID-19 infection in section 8.3. Section 8.4 states the proposed COVID-19 detection methodology. In section 8.5, a brief description of ANNs is mentioned. Sections 8.6 and 8.7 display the prediction of COVID-19 infection risk using ANN and experimental results and discussion, respectively, followed by the performance comparison between ANN and other classification algorithms in section 8.8. Lastly, we present some concluding comments.

8.2 Related studies

ANNs have found applications in many disciplines that include system identification, pattern recognition, medical diagnosis, social network filtering, and much more. A brief review is presented here.

Jiang, Trundle, and Ren (2010) have used the ANN for resolving medical imaging problems with fixed structure and training procedures. They have analyzed, processed, and characterized medical images and resolved problems relevant to medical imaging by expanding neural networks.

Huang, Shen, and Duong (2010) developed a flexible ANN to predict ischemic tissue fate and permanent cerebral artery occlusion in rats. To improve prediction accuracy neighboring pixel information and infarction incidence were included in the ANN model. The major finding was the objective framework that the ANN predictive model can provide to evaluate stroke treatment on an individual patient basis.
Sinha and Wang (2008) developed the ANN prediction model to predict the values of maximum dry density, permeability, and optimum moisture content with classification properties of the soils. The model gave an accuracy of 95%.

Hsu, Gupta, and Sorooshian (1995) presented an effective alternative to the autoregressive moving average with exogenous inputs (ARMAX) time series approach using a new procedure for the ANN model for developing input–output simulation and forecasting models. This new model will be able to simulate the nonlinear hydrologic behavior of watersheds in situations that do not require modeling of the internal structure of the watershed.

Hill, Marquez, O’Connor, and Remus (1994) presented a thorough comparison between ANNs and statistical models. The topic of comparison was based on regression-based forecasting, time series forecasting, and decision making. The primary objective of this report is to provide an assessment of ANN for forecasting and decision-making models.

Lee, Cha, and Park (1992) applied an ANN to forecast the short-term load for a large power system. Various combinations of the neurons with one or two hidden layers were tested. The predicted results were then compared in terms of forecasting error. The conclusion of this chapter claims that a good load forecast is obtained when the neural networks are grouped into different load patterns.

Kumar et al. (2002) estimated the daily grass reference crop evapotranspiration (ETo) using ANNs. Based on the weighted standard error of estimate and minimal ANN architecture, the best ANN architecture was selected. The results were compared with the conventional methods (Penman – Monteith) and found to be much more satisfactory by using ANN to predict ETo.

William G. Baxt (Baxt, 1991) used an ANN to identify myocardial infarction in patients. The network was tested on 331 consecutive patients with anterior chest pain. The network was compared with that of physicians caring for the same patient. A better result was obtained by the ANN.

Guresen, Kayakutlu, and Daim (2011) evaluated the effectiveness of neural network models. The models that are analyzed include multilayer perceptron, hybrid neural network, and dynamic ANN. The models are then compared based on Mean Square Error and Mean Absolute Deviate. Kumari et al. (2020), Bhatnagar et al. (2020), and Singh et al. (2020) analyzed the current situation and forecasted some about future scenario.

8.3 Fundamental symptoms and conditions responsible for COVID-19 infection

During the incubation period of 14 days, some infected people show no symptoms while others show a wide range of symptoms. The two most common symptoms include fever (88%) and dry cough (68%). The less common
symptoms include fatigue (38%), shortness of breath (19%), sputum production (33.4%), persistent chest pain, headache (14%), sore throat (13.9%), chills (11%), nasal congestion (4.8%), nausea (5%), diarrhea (4%), haemoptysis (0.9%), and pink eyes and lips (0.8%) (Huang et al., 2020; WHO – China Joint Mission, 2020; World Health Organization, 2020a). Though the recovery rate of the patients increased in the due course of time, the fatality rate of patients with cardiovascular disease and hypertension (Huang et al., 2020) has remained constant. The majority of the patients who suffered death due to COVID-19 progressed to the critical condition, which includes respiratory failure, septic shock, multiple organ failure (WHO – China Joint Mission, 2020; World Health Organization, 2020a; Chen et al., 2020; World Health Organization, 2020b). Patients with mild symptoms like common cold and cough tend to recover within 2 weeks (Zu et al., 2020). However, those patients suffering from severe symptoms like high respiratory frequency or dyspnea, in best cases recover within a span of 3 – 6 weeks.

The transmission rate of coronavirus is higher than that of influenza. Patients showing mild or nonspecific symptoms are highly infectious. Such patients remain infectious for an average of 2 weeks and can transmit the virus even from a single respiratory droplet. Studies have shown that people without symptoms also transmit the virus, which is known as asymptomatic transmission (U.S. Centers for Disease Control and Prevention CDC, 2020). The World Health Organization recommends 1 meter of social distance to restrict the primary spread of the virus from close contacts and from inhaling small droplets produced by an infected person by coughing or sneezing (Wang, Tang, & Wei, 2020; Rothen and Byrareddy, 2020; Zheng, Ma, Zhang, & Xie, 2020; Fang, Karakiulakis, & Roth, 2020; Guan et al., 2020; The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases COVID-19, 2020). The other modes of transmission include the suspension of the droplet in the air for quite a long period. The virus can also spread from the contaminated droplets that have fallen to floors or surfaces. Though the level of contamination that can transmit the infection via surfaces is not known, proper surface disinfectants should be used to restrict the unbounded spread of the virus.

### 8.4 Proposed COVID-19 detection methodology

Since the outbreak of COVID-19, the virus has evolved at a tremendous rate. One of the major causes of the failure of COVID-19 detection (Biswas, Sharma, Ranjan, & Banerjee, 2020; Guhathakurata, Kundu, Chakraborty, & Banerjee, 2020; Guhathakurata, Saha, Kundu, Chakraborty, & Banerjee, 2020a, 2020b) from the symptoms is mainly because of the uncertain nature of the virus. As a result of which no proper dataset is available to use as a reference. A synthetic multicriterion (Banerjee & Chakraborty, 2014, 2015;
Banerjee, Chakraborty, & Chattopadhyay, 2017, 2018b, 2018c, 2018a, 2021; Banerjee, Chakraborty, & Karmakar, 2013; Banerjee, Goswami, & Nandi, 2014; Banerjee & Karmakar, 2012; Saha, Chakraborty, & Banerjee, 2019; Saha et al., 2017) dataset has been coined by the authors to serve the purpose of this chapter, which is mentioned in the appendix (Table A1). The attributes or symptoms and conditions that have been taken into consideration include age, body temperature, dry cough, chest pain, breathing rate, hypertension, cardiovascular diseases, diabetes, tiredness, current smoker, and contact with a person with fever or cold in the last few days. The attributes are discussed below concerning their impact related to COVID-19:

1. Age: The fatality rate of the old aged people is more compared to the young ones in the coronavirus cases. The main reason for this development is basically due to the strong immunity level in the younger persons compared to the aged ones.

2. Body temperature: One of the most significant COVID-19 symptoms is body temperature. Temperature above 100°F is mild fever and above 103°F is high fever. Body temperature can be easily detected by the thermal scanner using sensors (Das, Pandey, Chakraborty, & Banerjee, 2017; Paul, Chakraborty, & Banerjee, 2019; Roy, Dutta, Biswas, & Banerjee, 2020; Paul et al., 2017; Das, Pandey, & Banerjee, 2016).

3. Dry cough: Cough without mucus or phlegm is one of the common symptoms of COVID-19.

4. Tiredness: The early sign of COVID-19 infection.

5. Chest pain: Persistent chest pain with breathing problems gives rise to a critical condition.

6. Nasal congestion: One of the symptoms to be looked upon as an early warning sign of coronavirus.

7. Runny nose: This is quite a common symptom of cold but when coupled with fever and breathing problems the person can be predicted into the early stages of coronavirus.

8. Sore throat: Considered to be the early symptom of coronavirus. Patients should consult the doctor when suffering from a sore throat.

9. Diarrhea: Patients suffering from diarrhea are very prone to be getting infected by a coronavirus.

10. Breathing rate: Breathing rate above 30 breath/min is a critical condition.

11. Hypertension: Patient with stage 1 hypertension (140–159 blood pressure) are less susceptible to coronavirus compared to a patient with stage 2 hypertension (160 and above blood pressure).

12. Cardiovascular diseases: The fatality rate increases for patients with cardiovascular diseases.

13. Diabetes: One of the critical symptoms to look upon to check for the risk of coronavirus infection.
14. Current smoker: Current smokers are more susceptible to suffer from acute respiratory disease syndrome.

15. Contact with a person with fever and cold in the last few days: This is a very important condition to be checked upon to track the source of coronavirus.

In this chapter, the authors have not considered some important factors like loss of smell and taste because nowadays it is almost the proven fact that loss of smell and taste is due to COVID-19. If anyone is having a loss of smell and taste, then he/she is advised to test for COVID-19. In this chapter, the authors have considered mainly symptomatic patients. As it is really hard to predict asymptomatic patients, a bulk number of people are infected by COVID-19 asymptotically, and in due course, they also develop the antibody to resist COVID-19. In this way, herd immunity can be achieved through community transmission.

Based on these symptoms and conditions, the risk condition of the patient can be deduced using our proposed approach. The risk or the infection status has been classified into three classes, which are Not Infected, Mildly Infected, and Severely Infected. The classes have been mapped to the words in the “RISK” column of the dataset as:

Not Infected = Low; Mildly Infected = Medium; Severely Infected = High

**Case 1:** Not Infected

Among all the symptoms that have been considered, certain symptoms occur quite often for any human being. For example, people with a common cold also experience a dry cough, but only these conditions are not enough to ascertain COVID-19 infection. Moreover, billions of people fall under the category of the current smoker, but without the presence of the other symptoms we can easily declare the person to be in the class of Not Infected.

**Case 2:** Mildly Infected

A person showcasing one or two symptoms simultaneously cannot be confirmed with COVID-19 but can progress to a critical stage if proper containment measures are not taken. Such a person falls under this class of mildly infected.

**Case 3:** Severely Infected

A person with more than 3–4 symptoms, each of which has crossed their normal limits, shows positive results for COVID-19 in the majority of cases. A person can be judged to be at high risk if they are suffering from high fever, high breathing rates, and are facing persistent chest pain.
The dataset is then passed to the ANN classifier. Due to the ability to reproduce and model nonlinear processes, ANN has been used keeping in mind the inconsistency of the data related to coronavirus. In section 8.5, the basic principles of ANN are described along with the proper justification of using this model in our proposed approach. Finally, ANN predicts the output among one of the three classes, that is, Not Infected, Mildly Infected, and Severely Infected (see Fig. 8.1).

8.5 Brief description of artificial neural networks

ANN, an integral part of machine learning (Biswas et al., 2021; Saha, Saha, Kundu, Chakraborty, & Banerjee, 2020; Banerjee et al., 2019; Pandey et al., 2019; Chattopadhyay et al., 2020), teaches the computer to think like a human. A computer can learn to execute classification jobs based on text, sound, or images. An ANN is trained by using a large dataset. Neural network architecture (Chakraborty & Banerjee, 2013; Chakraborty, Banerjee, & Chattopadhyay, 2017, 2019, 2020) contains many layers. ANN is an information processing system that is inspired by the way the nervous system processes information. It is formed of many highly interconnected processing elements (neurons) working together to solve a specific problem.

8.5.1 Principles of artificial neural network

To understand neural network, we need to discuss the following:

- Neurons—Biological neurons have inspired the development of the general model of ANN. Perceptron, which is a single layer of neural network, gives a single output.

In Fig. 8.2 for a single observation, $x_1, x_2, \ldots, x_n$ describes independent variables to the network. Each of these inputs are multiplied by connection weight $(w_1, w_2, \ldots, w_n)$. The strength of each node is depicted by the
respective weights. The activation function is shifted up or down through the bias value denoted by \( b \).

These products are primarily added and fed to an activation function to produce a result in the form of output that is sent.

\[
x_1 \cdot w_1 + x_2 \cdot w_2 + \ldots + x_n \cdot w_n = \sum_{i=1}^{n} x_i \cdot w_i
\]

The activation function is employed now, \( \varphi(\sum x_i \cdot w_i) \).

- **Activation functions**—An activation function is essential for the ANN. It helps to understand something very complicated. The main purpose is to process the input signals and convert them into the corresponding output signal. The next layer in the stack receives input from this output signal. The activation of the neuron depends upon the activation function. The decision is made by adding bias and calculating the weighted sum to it, which helps in the introduction of nonlinearity to the output signal coming out of a neuron. Without the activation function, the output signal would be a one-degree polynomial, that is, a linear function that is simple to solve but has the least power to solve complex problems.

Nonlinear functions have more than one degree and have a curvature.

“Universal function approximators,” is the other name of neural network, which means it tends to learn and compute any function.

Different kinds of activation functions:

1. **Binary step function or threshold activation function**—It is a threshold-based activation. If the input value is above a certain threshold and sends the same signal to the next layer, then only the neuron is activated (see Fig. 8.3).

   If \( Y > \text{threshold} \) then the activation function \( A = \text{“activated”} \) otherwise not or another way, if \( Y > \text{threshold 0} \), then \( A = 1 \).

2. **Logistic function or sigmoid activation function**—Sigmoid curve is a mathematical function with a characteristic “S”-shaped curve in the range of 0 and 1. Therefore it is primarily used to prognosticate the probability of an output.
As we can see in Fig. 8.4, we can find the curve’s slope at any two points.

We can infer from the above graph that:

- $F(z)$ tends toward 1 as $z \rightarrow \infty$
- $F(z)$ tends toward 0 as $z \rightarrow -\infty$
- $F(z)$ is always bounded between 0 and 1

3. Hyperbolic tangent function ($\tanh$)—It has similar characteristics to sigmoid, but it performs better. We can stack layers due to its nonlinear nature. The function ranges within $(-1, 1)$ (see Fig. 8.5).

This function maps the substantial negative inputs to negative outputs, and to near-zero outputs, only zero-valued inputs are mapped. Hence, during training, it is less likely to get stuck.

4. Rectified Linear Units (ReLu)—In conventional neural network and ANN, it is the most used activation function, ranging from zero to infinity.

If $x$ is positive, then ReLu gives an output “$x$” and 0 otherwise. A combination of ReLu is also nonlinear as ReLu is nonlinear. Any function can be approximated with a combination of ReLu as it is a valid approximator (see Fig. 8.6).
To the hidden layers of neural network, ReLu can only be applied. As the problem we are solving is a classification problem, we need to use a SoftMax function for the outer layer.

How do neural networks work—The input values pass within weighted synapses to the output layer. All the input parameters are to be analyzed, an activation function is employed to it in the neuron, and then the result will be produced.

There is a way to improve the neural network’s power and increase its accuracy by adding hidden layers that sit between output and input layers.

- All the four variables are attached to neurons through a synapse; it is visible from Fig. 8.7. Still, all the synapses are not weighted; it is either a nonzero value or 0. The value 0 indicates that they will be rejected, and the nonzero value indicates importance. With different combinations of variables, many neurons do similar calculations. That is why neural network is so powerful.
In a very flexible way, the neurons interact and work, allowing each to look for specific things. This approach allows them to do a comprehensive search for what it is trained for.

- How do neural networks learn—The cost function plays a vital role in training neural nets. For all the neural network layers, to adjust the weight and threshold of the next input, the cost function is analyzed. The cost function is one-half of the squared difference actual and output value. The lower the cost function, the more accurate is the prediction. Thus after each run, the error function keeps on getting lesser and lesser. As long as there is a mismatch between the actual value and predicted value, the weights are continuously adjusted during each run. After this adjustment, the neural network is run again, which forms a new cost function. This process is repeated until we make the cost function small and acceptable for the application-specific task.

Fig. 8.8 shows how back-propagation works. It is applied continuously through a network until the error value is kept at a minimum.

### 8.6 Parameter settings for the proposed ANN model

The major aspect of using the ANN is because they can learn by themselves and give the output that is not limited to the input data provided to the neural
network. This network can learn from examples and apply them when a similar type of event arises. In the current situation of COVID-19 we need a model that can learn by itself.

For this study, we have split the dataset into 80% train sets and 20% test sets. The dataset contains numeric values for attributes like body temperature and breathing rate. Other than that, all other attributes have yes/no value. The age attribute has been categorized into three classes, which are mapped as A = 0–20 years, B = 21–40 years, and C > 41 years. The total number of inputs for each record is 15. As a result, the input layer of the network contains 15 nodes. This layer provides information from the outside world to the network. No computation is performed here. The output layer contains just a single node since we are predicting a single output among the three output classes. The number of nodes of the hidden layer is computed from the average of nodes of the input layer and output layer, which is 8. The hidden layers perform all the computations on the outside information brought by the input layer and transfer the result to the output layer.

An activation function is important for the ANN. It is responsible for making sense of something complicated. The main purpose is to process the input signal and convert them into the corresponding output signal. The purpose of the activation function is to introduce nonlinearity into the output of a neuron. The activation function calculates the weighted sum and decides whether a neuron should be activated or not and further adds bias with it. The purpose of the activation function is to introduce nonlinearity into the output of a neuron. We are using ReLu as the neural network’s activation function. ReLu is six times improved over hyperbolic tangent function. It is applied to the hidden layers of the neural network. It gives an output “x” if x is positive and 0 otherwise. ReLu is a good approximator and any function can be approximated with a combination of ReLu. Since the dataset created for this study consists of a minimal record, the batch size that has been used
for training the model has been kept as 10. As a result, for each iteration the algorithm takes 10 records to train the network and update the weights to minimize the error before proceeding with the next batch of 10 records. To improve the accuracy of the model the network has been trained for five iterations or epochs.

8.7 Experimental results and discussion

The dataset that has been created for this study contains 230 records, with 15 attributes that are described in section 8.3. Due to the unavailability of a proper dataset, authors are bound to employ the machine learning algorithm on a minimal number of data. The authors also believe that the proposed technique will perform better if provided with a real-time dataset. In this dataset (see Appendix, Table A1), the output column, that is, the RISK column, contains three different values, “Low,” “Medium,” and “High,” which have been mapped to Not Infected, Mildly Infected, and Severely Infected, respectively.

Fig. 8.9 gives a clear visualization of the risk case distribution of COVID-19 based on our dataset. The authors have been very meticulous in preparing the dataset where the impact of various factors and conditions have been taken into consideration to assign a case as high risk. As a result of which a better accuracy has been achieved while predicting a patient to be at high risk. Keeping in mind the importance of a more accurate prediction of COVID-19 risk, ReLu has been used as an activation function that has been discussed in section 8.5. The ANN has been implemented by using python in Spyder platform.

From the confusion matrix in Fig. 8.10, we see that the number of correct predictions is quite high, which is depicted by the diagonal values in the matrix. The total number of correct predictions is 39 (diagonal values) and the total number of records tested is 46. Therefore the accuracy becomes \( \frac{39}{46} \times 100 = 84.7 \) (approximately).

In Table 8.1, the classifier on the test set has shown high precision for Not Infected class and moderate results for the Severely Infected class. This means that very few cases have been labeled wrong. Since all the predictions of Mildly Infected cases were labeled wrong, the precision obtained for that class is 0. However, with real-time data, the model is expected to function more efficiently and thus provide more accurate results. The classifier has shown perfect results for predicting the cases that belong to the Severely Infected class since it has recall = 1. This is again a very crucial aspect since the correct prediction of high risk gives our approach more reliability. The f1-score is simply the average of precision and recall. The final column “Support” gives the number of true values for each class in the test data.
FIGURE 8.9  Distribution of risk cases of COVID-19.

FIGURE 8.10  Confusion matrix.

| Class           | Precision | Recall | f1-score | Support |
|-----------------|-----------|--------|----------|---------|
| Not Infected    | 1.00      | 0.30   | 0.47     | 33      |
| Mildly Infected | 0.00      | 0.00   | 0.00     | 3       |
| Severely Infected | 0.56    | 1.00   | 0.71     | 10      |

TABLE 8.1  Classification report.
8.8 Performance comparison between ANN and other classification algorithms

The authors in this chapter analyzed the COVID-19 dataset by predicting the risk factor using all the supervised models. Among all the other supervised models, ANN works best in predicting COVID-19 risk with maximum accuracy. To support our claim, we have performed a comparative study on the popular supervised learning models. Looking at the current pandemic scenario, it is visible how quickly the COVID-19 is spreading its roots and how its symptoms have evolved over the last few months since its outbreak. Therefore, the concrete data related to the symptoms will further increase the number of predictors for COVID-19 data analysis and predictions. Keeping all these in mind, we propose the usage of visual programming methodology using Orange (Maughan, 2019) to quickly analyze the data with a set of predictors and see their efficiency in predicting the disease.

Orange is an open source machine learning and data visualization toolkit. It not only facilitates easy analysis of data using machine learning models from researchers from any background but also reduces coding overheads. We have analyzed the performance comparison between different machine learning models using Orange. Figs. 8.11–8.13 depict the methodology implemented by Orange. Initially the COVID-19 dataset is parsed and based on the predictive score the features are ranked. This step ensures the selection of features that have the best correlation with the predictor. In general, the number of features gets reduced. Moreover, this “Rank” block helps us to identify the linearly varying redundant features and thus the overall computation time gets reduced (see Fig. 8.13). This phenomenon makes the methodology fast, scalable, and robust. Popular scoring methods (Orange Visual Programming, 2020) like Info Gain, FCBF, Gini, and ReliefF are employed for ranking purposes. After that the top features are selected and fed into the machine learning models for training and prediction. The results of the prediction are evaluated by the “Test and Score” module and also provides comparison metrics. These metrics give us a clear visualization of the

![Evaluation Results](image)

FIGURE 8.11 Evaluation results.
performance of various supervised machine learning models viz. Tree, KNN, AdaBoost, Random Forest, Neural Network, and Logistic Regression. The evaluation results of the models are depicted in Fig. 8.10 and ANN outperforms all other models that are tested. We have done our prediction based on the parameters like Area under receiver operating characteristic (ROC), Classification Accuracy, F1 Score, Precision, and Recall. The confusion

**FIGURE 8.12** Schematics of visual programming.

**FIGURE 8.13** Rank of features.
matrix for all the models has been summarized in Fig. 8.14, illustrating the superiority of ANN in predicting COVID-19 efficiently.

8.9 Conclusion

This chapter aims to develop a model that can predict whether a person is affected by COVID-19 or not using ANN. The symptoms or attributes and conditions have been selected for the dataset, keeping in mind the evolving nature of the virus and its uncertainty. The conditions considered as an attribute in the dataset are to give more accuracy in predicting and segregating the Not Infected class and the Mildly Infected class. The Not Infected class, for the time being, can be considered safe and cannot be said with certainty since persons without any signs and symptoms of COVID-19 are also getting affected. A person who has been predicted to be in the Mildly Infected class should undergo proper medication and containment measures to prevent this condition’s elevation to a severe state. A person who is in the Severely Infected class should immediately undergo proper medication to overcome the disease. Authors have shown that ANN outperforms other supervised models for the purpose of predicting COVID-19. The accuracy obtained by implementing ANN by Python in the Spyder platform is 84.7% and in Orange platform is 99.1%. Due to the minimal size of the dataset and different platforms, the accuracy obtained is different in both cases. To reduce the margin of error concerning the platform predicting the result, the authors have considered 84.7% accuracy to predict COVID-19. Moreover, with each new entry of the data related to COVID-19 patients, the prediction accuracy improves. Such type of approach improves the results in better prediction and on-time treatment of the patient.
## Appendix

### TABLE A1 Dataset snapshot.

| Age  | Body temp | Dry cough | Tiredness | Chest pain | Nasal congestion | Runny nose | Sore throat | Diarrhea | Breathing | Hypertension | Cardiovascular Diseases | Diabetes | Current smoker | Contact with a person with fever | Risk |
|------|-----------|-----------|-----------|------------|------------------|------------|-------------|----------|-----------|--------------|--------------------------|-----------|-----------------|---------------------------------|------|
| A    | 98.7      | Yes       | No        | No         | Yes              | No         | Yes          | No       | 15        | No           | Yes                      | No        | No              | Yes                             | Low  |
| A    | 100.2     | No        | No        | No         | No               | No         | No           | No       | 16        | No           | No                      | No        | No              | No                              | Low  |
| A    | 102.8     | Yes       | Yes       | No         | No               | Yes        | No           | No       | 14        | No           | No                      | No        | No              | No                              | Low  |
| A    | 99        | No        | No        | No         | No               | No         | No           | No       | 18        | No           | No                      | No        | No              | No                              | Low  |
| A    | 99.6      | Yes       | Yes       | No         | No               | No         | Yes          | 17       | No        | No           | No                      | No        | No              | No                              | Low  |
| A    | 101.1     | No        | No        | No         | No               | No         | No           | 16       | No        | No           | No                      | No        | No              | No                              | Low  |
| A    | 100.8     | No        | No        | No         | No               | No         | No           | No       | 20        | Yes          | Yes                     | No        | No              | No                              | Low  |
| A    | 99.2      | Yes       | No        | No         | Yes              | No         | No           | No       | 25        | No           | No                      | No        | No              | No                              | Low  |
| A    | 99.5      | Yes       | Yes       | No         | No               | No         | No           | 14       | No        | No           | No                      | No        | No              | No                              | Low  |
| A    | 98.4      | No        | No        | No         | No               | No         | No           | 16       | No        | No           | No                      | No        | No              | No                              | Low  |
| A    | 101.3     | No        | No        | No         | No               | No         | Yes          | Yes      | 18        | No           | No                      | No        | No              | No                              | Low  |
| A    | 99        | Yes       | Yes       | No         | Yes              | No         | No           | No       | 17        | No           | No                      | No        | No              | No                              | Low  |
| A    | 100.6     | No        | No        | No         | No               | No         | No           | 15       | No        | No           | No                      | No        | No              | No                              | Low  |
| A    | 98.1      | Yes       | No        | No         | No               | No         | Yes          | 13       | No        | No           | No                      | No        | No              | No                              | Low  |
| A    | 103       | No        | Yes       | No         | No               | No         | No           | 16       | No        | No           | No                      | No        | No              | No                              | Low  |
| A    | 99.7      | Yes       | No        | No         | No               | No         | No           | 18       | No        | No           | No                      | No        | No              | No                              | Low  |
| A    | 102.2     | No        | No        | No         | No               | No         | No           | 16       | No        | No           | No                      | No        | No              | No                              | Low  |
| A    | 98.9      | Yes       | No        | No         | Yes              | No         | No           | 24       | No        | No           | No                      | No        | No              | No                              | Low  |
| A    | 103.5     | Yes       | Yes       | Yes        | No               | Yes        | Yes          | 28       | No        | No           | No                      | No        | No              | Yes                             | Medium |
References

Banerjee, J. S., & Chakraborty, A. (2014). Modeling of Software Defined Radio Architecture and Cognitive Radio: The Next Generation Dynamic and Smart Spectrum Access Technology. In Cognitive radio sensor networks: Applications, architectures, and challenges (pp. 127–158). IGI Global.

Banerjee, J. S., & Chakraborty, A. (2015). Fundamentals of software defined radio and cooperative spectrum sensing: a step ahead of cognitive radio networks. In Handbook of research on software-defined and cognitive radio technologies for dynamic spectrum management (pp. 499–543). IGI Global.

Banerjee, J., Maiti, S., Chakraborty, S., Dutta, S., Chakraborty, A., & Banerjee, J.S. (2019). Impact of machine learning in various network security applications. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), (pp. 276–281). IEEE.

Banerjee, J. S., Chakraborty, A., & Chattopadhyay, A. (2017). Fuzzy based relay selection for secondary transmission in cooperative cognitive radio networks. Advances in Optical Science and Engineering (pp. 279–287). Singapore: Springer.

Banerjee, J. S., Chakraborty, A., & Chattopadhyay, A. (2018a). A novel best relay selection protocol for cooperative cognitive radio systems using fuzzy AHP. Journal of Mechanics of Continua and Mathematical Sciences, 13(2), 72–87.

Banerjee, J. S., Chakraborty, A., & Chattopadhyay, A. (2018b). Relay node selection using analytical hierarchy process (AHP) for secondary transmission in multi-user cooperative cognitive radio systems. Advances in Electronics, Communication and Computing (pp. 745–754). Singapore: Springer.

Banerjee, J. S., Chakraborty, A., & Chattopadhyay, A. (2018c). Reliable best-relay selection for secondary transmission in co-operation based cognitive radio systems: A multi-criteria approach. Journal of Mechanics of Continua and Mathematical Sciences, 13(2), 24–42.

Banerjee, J. S., Chakraborty, A., & Chattopadhyay, A. (2021). A decision model for selecting best reliable relay queue for cooperative relaying in cooperative cognitive radio networks: the extent analysis based fuzzy AHP solution. Wireless Networks, 27(4), 2909–2930.

Banerjee, J. S., Chakraborty, A., & Karmakar, K. (2013). Architecture of cognitive radio networks. In Cognitive radio technology applications for wireless and mobile ad hoc networks, (pp. 125–152). IGI Global.

Banerjee, J. S., & Karmakar, K. (2012). A comparative study on cognitive radio implementation issues. International Journal of Computer Applications, 45(15), 44–51.

Banerjee, J. S., Goswami, D., & Nandi, S. (2014). OPNET: a new paradigm for simulation of advanced communication systems. In Proceedings of International Conference on Contemporary Challenges in Management, Technology & Social Sciences, SEMS (pp. 319–328).

Baxt, W. G. (1991). Use of an artificial neural network for the diagnosis of myocardial infarction. Annals of Internal Medicine, 115(11), 843–848.

Bhattacharyya, V., Poonia, R. C., Nagar, P., Kumar, S., Singh, V., Raja, L., & Dass, P. (2020). Descriptive analysis of COVID-19 patients in the context of India. Journal of Interdisciplinary Mathematics, 1–16. Available from https://doi.org/10.1080/09720502020.2020.1761635.

Biswas, S., Sharma, L. K., Ranjan, R., Saha, S., Chakraborty, A., & Banerjee, J. S. (2021). Smart farming & water saving based intelligent irrigation system implementation using the internet of things. In Recent Trends in Computational Intelligence Enabled Research (pp. 339–354). Academic Press, Elsevier.
Biswas, S., Sharma, L. K., Ranjan, R., & Banerjee, J. S. (2020). Go-COVID: An interactive crossplatform based dashboard for real-time tracking of COVID-19 using data analytics. *Journal of Mechanics of Continua and Mathematical Sciences, 15*(6), 1–15.

Chattopadhyay, J., Kundu, S., Chakraborty, A., & Banerjee, J. S. (2020). Facial expression recognition for human computer interaction. In *International Conference on Computational Vision and Bio Inspired Computing*, (pp. 1181–1192). Springer, Cham.

Chakraborty, A., & Banerjee, J. S. (2013). An advance Q learning (AQL) approach for path planning and obstacle avoidance of a mobile robot. *International Journal of Intelligent Mechatronics and Robotics (IJIMR), 3*(1), 53–73.

Chakraborty, A., Banerjee, J. S., & Chattopadhyay, A. (2017). Non-uniform quantized data fusion rule alleviating control channel overhead for cooperative spectrum sensing in cognitive radio networks. In *2017 IEEE 7th International Advance Computing Conference (IACC)* (pp. 210–215). IEEE.

Chakraborty, A., Banerjee, J. S., & Chattopadhyay, A. (2019). Non-uniform quantized data fusion rule for rate saving and reducing control channel overhead for cooperative spectrum sensing in cognitive radio networks. *Wireless Personal Communications, 104*(2), 837–851.

Chakraborty, A., Banerjee, J. S., & Chattopadhyay, A. (2020). Malicious node restricted quantized data fusion scheme for trustworthy spectrum sensing in cognitive radio networks. *J. Mech. Contin. Math. Sci, 15*(1), 39–56.

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., & Yu, T. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *The Lancet*, 395(10223), 507–513.

Das, D., Pandey, I., Chakraborty, A., & Banerjee, J. S. (2017). Analysis of implementation factors of 3D printer: the key enabling technology for making prototypes of the engineering design and manufacturing. *International Journal of Computer Applications, 975*, 8887.

Das, D., Pandey, I., & Banerjee, J. S. (2016). An in-depth study of implementation issues of 3D printer. In *Proceedings of MICRO 2016 Conference on Microelectronics, Circuits and Systems* (pp. 45-49).

Fang, L., Karakiulakis, G., & Roth, M. (2020). Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *The Lancet. Respiratory Medicine*.

Guan, W. J., Ni, Z. Y., Hu, Y., Liang, W. H., Ou, C. Q., He, J. X., & Du, B. (2020). Clinical characteristics of coronavirus disease 2019 in China. *New England Journal of Medicine, 382*(18), 1708–1720.

Guhathakurata, S., Kundu, S., Chakraborty, A., & Banerjee, J. S. (2020). A novel approach to predict COVID-19 using support vector machine. In *Data Science for COVID-19*, (pp. 351–364). Elsevier.

Guhathakurata, S., Saha, S., Kundu, S., Chakraborty, A., & Banerjee, J. S. (2020a). *South Asian countries are less fatal concerning COVID-19: A hybrid approach using machine learning and M-AHP. Computational Intelligence Techniques for Combating COVID-19*. Springer (press).

Guhathakurata, S., Saha, S., Kundu, S., Chakraborty, A., & Banerjee, J. S. (2020b). South Asian countries are less fatal concerning COVID-19: A fact-finding procedure integrating machine learning & Multiple Criteria Decision Making (MCDM) technique. *Journal of the Institution of Engineers (India): Series B*, Springer (Communicated).

Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications, 38*(8), 10389–10397.

Hill, T., Marquez, L., O’Connor, M., & Remus, W. (1994). Artificial neural network models for forecasting and decision making. *International Journal of Forecasting, 10*(1), 5–15.
Hsu, K. L., Gupta, H. V., & Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process. *Water Resources Research, 31*(10), 2517–2530.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., & Cheng, Z. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet, 395*(10223), 497–506.

Huang, S., Shen, Q., & Duong, T. Q. (2010). Artificial neural network prediction of ischemic tissue fate in acute stroke imaging. *Journal of Cerebral Blood Flow & Metabolism, 30*(9), 1661–1670.

Jiang, J., Trundle, P., & Ren, J. (2010). Medical image analysis with artificial neural networks. *Computerized Medical Imaging and Graphics, 34*(8), 617–631.

Kumar, M., Raghuvanshi, N. S., Singh, R., Wallender, W. W., & Pruitt, W. O. (2002). Estimating evapotranspiration using artificial neural network. *Journal of Irrigation and Drainage Engineering, 128*(4), 224–233.

Kumari, R., Kumar, S., Poonia, R. C., Singh, V., Raja, L., Bhatnagar, V., & Agarwal, P. (2020). Analysis and predictions of spread, recovery, and death caused by COVID-19 in India. *Big Data Mining and Analytics, IEEE*. Available from https://doi.org/10.26599/BDMA.2020.9020013.

Lee, K. Y., Cha, Y. T., & Park, J. H. (1992). Short-term load forecasting using an artificial neural network. *IEEE Transactions on Power Systems, 7*(1), 124–132.

Maughan, J. Machine Learning with Orange, (2019). https://medium.com/@jackmaughan_50251/machine-learning-with-orange-8bc1a541a1d7/ (accessed August 10, 2020).

Novel, C.P.E.R.E. (2020). The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi, 41*(2), 145.

Orange Visual Programming (2020). https://orange-visual-programming.readthedocs.io/widgets/data/rank.html/ (accessed August 10, 2020).

Pandey, I., Dutta, H.S., & Banerjee, J.S. (2019). WBAN: A smart approach to next generation e-healthcare system. In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, (pp. 344–349). IEEE.

Paul, S., Chakraborty, A., & Banerjee, J.S. (2017). A fuzzy AHP-based relay node selection protocol for wireless body area networks (WBAN). In *2017 4th International Conference on Opto-Electronics and Applied Optics (Optronix)*, (pp. 1–6). IEEE.

Paul, S., Chakraborty, A., & Banerjee, J. S. (2019). *The extent analysis based fuzzy AHP approach for relay selection in WBAN*. Cognitive Informatics and Soft Computing (pp. 331–341). Singapore: Springer.

Rothan, H. A., & Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of Autoimmunity, 102*433.

Roy, R., Dutta, S., Biswas, S., & Banerjee, J.S., *Android things: A comprehensive solution from things to smart display and speaker*. In *Proceedings of International Conference on IoT Inclusive Life (ICIIL 2019), NITTTRChandigarh, India* Singapore (pp. 339–352) Springer.

Saha, O., Chakraborty, A., & Banerjee, J.S. (2017). A decision framework of IT-based stream selection using analytical hierarchy process (AHP) for admission in technical institutions. In *2017 4th International Conference on Opto-electronics and Applied Optics (Optronix)*, (pp. 1–6). IEEE.

Saha, O., Chakraborty, A., & Banerjee, J. S. (2019). A fuzzy AHP approach to IT-based stream selection for admission in technical institutions in India. *Emerging Technologies in Data Mining and Information Security* (pp. 847–858). Singapore: Springer.

Saha, P., Saha, S., Kundu, S., Chakraborty, A., & Banerjee, J. S. (2020). Application of machine learning in app-based cab booking system: A survey on Indian scenario. *Applications of Artificial Intelligence in Engineering*. Springer (Press).
Singh, V., Poonia, R. C., Kumar, S., Dass, P., Agarwal, P., Bhatnagar, V., & Raja, L. (2020). Prediction of COVID-19 coronavirus pandemic based on time series data using support vector machine. *Journal of Discrete Mathematical Sciences & Cryptography*. Available from https://doi.org/10.1080/09720529.2020.1784525.

Sinha, S. K., & Wang, M. C. (2008). Artificial neural network prediction models for soil compaction and permeability. *Geotechnical and Geological Engineering*, 26(1), 47–64.

U.S. Centers for Disease Control and Prevention (CDC), (2020). How COVID-19 spreads. Retrieved August 6, 2020.

United States Centers for Disease Control and Prevention (CDC), (2020a). Interim Clinical Guidance for Management of Patients with Confirmed Coronavirus Disease (COVID-19). August 4, 2020.

United States Centers for Disease Control and Prevention (CDC), (2020b). Symptoms of Novel Coronavirus (2019-nCoV). Retrieved February 10, 2020.

Velavan, T. P., & Meyer, C. G. (2020). The COVID-19 epidemic. *Tropical Medicine & International Health*, 25(3), 278–280. Available from https://doi.org/10.1111/tmi.13383. PMC7169770, PMID 32052514.

Wang, W., Tang, J., & Wei, F. (2020). Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. *Journal of Medical Virology*, 92(4), 441–447. Webmd, (2020). https://www.webmd.com/lung/covid-19-symptoms#1 (Retrieved on: September 23, 2020).

WHO – China Joint Mission, (2020). Report of the WHO – China Joint Mission on Coronavirus Disease 2019 (COVID-19). https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-finalreport.pdf (accessed March 1, 2020).

Wikipedia, (2020). https://en.wikipedia.org/wiki/COVID-19_pandemic (Retrieved on: September 12, 2020).

World Health Organization, (2020a). Coronavirus disease 2019 (COVID-19) Situation Report—47. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200307-sitrep-47-covid19.pdf?sfvrsn=27c364a4_2 (accessed March 7, 2020).

World Health Organization, (2020b). Coronavirus disease 2019 (COVID-19) Situation Report—47. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200307-sitrep-47-covid19.pdf?sfvrsn=27c364a4_2 (accessed August 6, 2020).

Zheng, Y. Y., Ma, Y. T., Zhang, J. Y., & Xie, X. (2020). COVID-19 and the cardiovascular system. *Nature Reviews Cardiology*, 17(5), 259–260.

Zu, Z. Y., Jiang, M. D., Xu, P. P., Chen, W., Ni, Q. Q., Lu, G. M., & Zhang, L. J. (2020). Coronavirus disease 2019 (COVID-19): A perspective from China. *Radiology*, 200490.