A Comparative Study Based on Deep Learning and Machine Learning Methods for COVID-19 Detection Using Audio Signal

Fulya Akdeniz[(✉)], Merve Nur Damar, Buse İrem Danacı, Burcu Kır Savaş,
and Yaşar Becerikli

Kocaeli University, Umuttepe Campus, Kocaeli 41001, Turkey
{fulya.akdeniz,200202091,190201123,burcu.kir,
ybecerikli}@kocaeli.edu.tr

Abstract. COVID-19 is an upper respiratory disease that emerged in the last months of 2019 and affecting people worldwide. According to the World Health Organization data, it is determined that the outbreak spread rapidly. Early diagnosis is very important in preventing the spread of epidemics. Highly accurate identification and isolation of infected persons is essential to prevent the spread of outbreaks. Currently, reverse transcription polymerase chain reaction (RT-PCR) test is the most used method to detect COVID-19. However, the test is expensive, time-consuming, socially distant, difficult to distribute and has a high false negative rate. In this study, we developed a system based on Machine Learning (ML) and Deep Learning (DL) methods using voice signals (cough sounds) for early detection and diagnosis of COVID-19 disease. Since cough sounds are also a symptom of many upper respiratory tract diseases, a three-class system was developed to determine whether the voice recording from a person was COVID-19, upper respiratory tract disease, healthy or not. In this paper, indicative features were extracted Mel Frequency Cepstrum Coefficients (MFCC) from audio signals and features were classified with K-Nearest Neighbor Algorithm (KNN), Random Forest (RF), Decision Tree, Support Vector Machines (SVM), Logistic Regression classifiers. In addition, Convolutional Neural Networks (CNN) methods, one of the DL methods, was used and comparative performance results are given in the experimental section.

Keywords: Deep learning · Audio signal · Covid-19 detection · MFCC

1 Introduction

Coronavirus Disease 2019, COVID-19 is an infectious disease caused by a coronavirus called SARS-CoV-2 [1]. Early detection of COVID-19 plays an important role in reducing the rate of spread of the disease and the damage caused by providing rapid decision-making [2].

The RT-PCR test used to diagnose COVID-19 disease is a very costly test and the time taken to obtain the result is quite long. Many studies emphasize that the false negative rates in the results of this test are also quite high. In other words, false negative results are

quite high in people with COVID -19 virus levels [3]. X-Ray and Computed Tomography (CT) images are frequently used to support PCR tests to diagnose COVID -19 disease. However, diseases symptoms such as COVID-19, pneumonia, bacterial pneumonia on the lung are quite similar. Therefore, it is very difficult to distinguish between these diseases on X-Ray [4]. Chest X-Ray images of COVID -19 patients are analyzed, COVID -19 findings on X-Ray peaked 10–12 days after the onset of symptoms of the disease. CT images are still inadequate to identify and differentiate specific viruses [5]. X-rays can be harmful because they are classified as ionizing radiation [6, 7]. Furthermore, exposure of pregnant and lactating women and children to X-rays is harmful. In this study, Covid-19 disease is detected using low-cost, easy-to-use audio signals that are harmless to human health. Most studies show that COVID-19 disease severely affects human breathing, audio and causes symptoms that make voices of patients distinctive [8, 9]. Once a person's audio recordings can be used to detect COVID-19 disease after are analyzed. It is very important to use audio signals to diagnose COVID-19 since audio signals are easy to obtain, can be analyzed quickly and effectively, and are cost-effective. Differentiating the pandemic-causing COVID-19 virus from other less dangerous upper respiratory tract diseases is an important factor during the pandemic [10]. Coughing is among the most common and most important symptoms of COVID-19. However, cough sounds are related to many diseases [11]. Therefore, it is important to differentiate COVID-19 from upper respiratory tract diseases. Differentiating between COVID-19 and other symptomatic diseases is important for accurate diagnosis and correct treatment. The symptoms of COVID-19 can be similar to some other symptomatic diseases. However, since COVID -19 has a high contagiousness and spread rate, it is important to diagnose the disease correctly to reduce the risk of transmission.

In this study, Mel frequency-based features were extracted from cough sounds. In the first stage, the features were classified into three classes using ML methods. In the second stage, the features were classified into three classes with DL architecture and different hyperparameters were used to determine whether the individual is healthy, COVID-19 and symptomatic. The contributions of this work are as follows:

- A three-class detection and classification system were developed that automatically detected COVID-19 disease using cough sounds.
- ML and DL algorithms optimized the hypermeters of different models for the audio-based model.

2   Related Works

Dash et al. found that cough sounds used for COVID-19 detection have different frequencies than speech sounds and developed a COVID-19 Coefficient feature by optimizing the frequency range and frequency conversion scale of the MFCC method to improve the efficiency of COVID-19 detection in their study. They used Coswara and Crowdsourced Respiratory Sound Data datasets. They tested various feature groups using a simple SVM-based classifier [12]. Grant et al. used breath, cough and speech sounds for Covid-19 detection in their study. They extracted MFCC, delta-MFCC and RASTA-PLP features from the signal. They obtained the AUC performance result of the study as 79.38% [13]. Despotovic et al. used a dataset of breath, cough and voice recordings collected from COVID-19 infected and uninfected people. A total of 1103 data, 84

COVID-19 data and 1019 healthy data from their dataset, were used in the study. They obtained Geneva Minimalistic Acoustic Parameter Set (GeMaps), extended GeMaps and ComParE feature set features from the signal. In the classification phase, they created Random Forests, VGGish and OpenL3 algorithms and achieved a performance result of 88.52% [8]. Coppock et al. detected symptomatic and asymptomatic COVID-19 cases using breath and cough voice recordings. The CNN model was tested on 355 data in the study. They then achieved an AUC rate of 84.6% [14]. Chaudhari et al., detected COVID-19 disease in crowd-sourced cough sound recorded and collected on smartphones from various regions of the world. They used MFCC and Mel-spectrogram features on developed CNN model. A total of 2883 data from Coswara and Coughvid datasets, 539 COVID-19 data and 2344 healthy data, were used. Then, their performance result is obtained with a ROC-AUC value of 77.1% [15]. Tena et al. extracted Autoencoder time-frequency features from cough sounds to detect COVID-19 disease. They used various machine learning algorithms (RF, SVM, LR, NB, LDA) in the classification phase. Coswara, Virufy and the Pertussis datasets that collected a total of 813 data were used in the study. They obtained the best accuracy rate is 90% from the RF algorithm [16]. Han et al. generated a total of 828 audio samples from 343 patients by inputting audio signals and symptoms of the patients. The features were extracted with Zero Crossing Rate (ZCR), Root Mean Square (RMS) frame energy, pitch frequency (F0), Harmonicsto-Noise Ratio, and MFCCs. They used SVMs algorithm in the classification stage. Their performance results are obtained with AUC 79%, sensitivity 68% and specificity 82% [17]. Alsabek et al. compared the success of cough, respiration and speech sounds in COVID-19 detection. The study is examined the importance of speech signal processing in MFCC extraction of COVID-19 and non-COVID-19 samples and the relationship between these samples using Pearson correlation coefficients (PCCs). A total of 42 data were used in this study, including 3 recordings from 7 COVID-19 patients and 3 recordings from 7 negative patients. They also used CNN and LSTM algorithms were used in their studies, too [18]. Çelik et al. proposed a DL-based Covid-CoughNet method for COVID-19 detection using cough, respiration and voice signals. He used Chroma features, RMS Energy, Spectral centroid, Spectral bandwidth, Spectral rolloff, ZCR, and MFCC methods for feature extraction. InceptionFireNet algorithm was used in the classification stage. As a result of his studies on two different datasets; In the COUGHVID dataset, the AUC value was measured as 98.44% when classified as healthy, COVID-19 and symptomatic. In the use of voice Coswara dataset gave an AUC value of 99.24% [19]. Ulukaya et al. proposed a deep neural network-based model for detect coronavirus from coughing sound only, fast, can be used remotely and has no harmful side effects. The proposed model named MSCCov19Net uses MFCC, Spectrogram and Chromagram features as an input. The system was trained on the Coswara, Coughvid, Virufy and NoCoCo-Da datasets and tested on two unseen (used only for testing) clinical and non-clinical datasets. The performance of MSCCov19Net is compared with various CNN architectures. The proposed system achieved 61.5% accuracy on Virufy and 90.4% accuracy on NoCoCoDa [20]. Kranthi Kumar et al. propose a lightweight CNN with Modified-MFCC using different parameters to classify COVID-19 and other respiratory voice disease symptoms. They used the CU dataset in their study. The proposed model outperforms traditional feature extraction models and existing DL

models in the range of 4% to 10% in COVID-19/SARS-CoV-2 classification with an accuracy of 93.65% [21].

3   Methods and Database

3.1. MFCCs: MFCCs are a measure of the human ear perceives a change in frequency. The human ear is generally sensitive to lower frequencies. With MFCC, this perception of the human ear is visualized [22]. MFCC is a widely used method for representing audio signals as feature vectors. It is developed to mimic human sound perception. MFCC-based features are proposed to minimize the difference and avoid perceptual and feature space distortion [12, 23].

3.2. SVM: It is a classification method used in machine learning. It uses the attributes in the dataset and generates a classification and regression model based on these attributes. An SVM has a great generalization capability, considering account errors [24]. SVM advantages are it works well on multidimensional data and reduces overfitting, is less sensitive to discrete values and noise in the data set and can separate data sets that cannot be linearly separated by transforming them into spaces. The parameters used in the SVM algorithm are LINEAR, Radial Basis Function-RBF, sigmoid or precomputed kernels, which are used to determine how the data points are separated in space and the decision boundary.

3.3. KNN: KNN is an algorithm used in machine learning and data mining for solving classification and regression problems [11]. KNN positions data points in a space and uses the labels of the nearest neighbors of these data points to predict data class or value of a new data point. The KNN algorithm uses different distance metrics to measure the difference between two samples [25].

3.4. Logistic Regression: Logistic Regression is a statistical computational method and machine learning algorithm for solving problems with two or more categorical classes. The class label of the dependent variable makes the probability estimate and the value of the dependent variable represents the class labels. Logistic Regression calculates probabilities using the sigmoid function [26]. Since it calculates probabilities, it estimates class labels in the interval [0, 1]. Multiple logistic regression is used in problems with more than one independent variable.

3.5. Decision Trees: Decision trees are among the first statistical algorithms for general-purpose prediction and classification mechanisms since the late 20th century [27]. Decision trees analyze a training set for which the class labels. Thus, they can classify previously unsupervised examples. With decision trees, high accuracy can be achieved on very high-quality data. A decision tree classifies or evaluates data samples using a set of decision rules or questions [28]. This algorithm's advantages are easy to understand, provide high performance and the results can be interpreted.

3.6. Random Forest: RF is a classification algorithm widely used ML. It is a combination of multiple decision trees. It is a powerful nonparametric statistical method that allows regression problems as well as two-class and multi-class classification problems to be addressed in a single and multifaceted framework [29].

3.7. CNN: CNN is an artificial neural network model frequently used in deep learning. CNN structure is based on neurons that are organized in layers. Neurons in these layers are connected to each other through weights and biases [30]. Its layers include convolution layer, pooling layer and fully connected layers etc. [31].

3.8. COUGHVID Database: Detecting various diseases from coughs is becoming more and more common today. One of the most common symptoms of COVID-19 is coughing. The COUGHVID dataset is one of the most widely used datasets in this field with more than 25000 cough records. In this dataset, all records were labeled by expert doctors and made suitable for classification. Of the labeled data, 25% received the Covid value, 35% received the symptomatic value, 25% received the healthy value and 15% received the condition-free value. The COUGHVID dataset was collected through a web application installed on a dedicated server at the École Polytechnique Fédérale de Lausanne (EPFL) in Switzerland. During these enrollments, information such as age, gender and geographical location is also collected from the user on request [32]. As such, COUGHVID is one of the largest and most reliable datasets.

4 Experimental Result

In this study, using ML and DL techniques is detected COVID-19 disease on the COUGHVID audio dataset. The results are comparatively analyzed in this section.
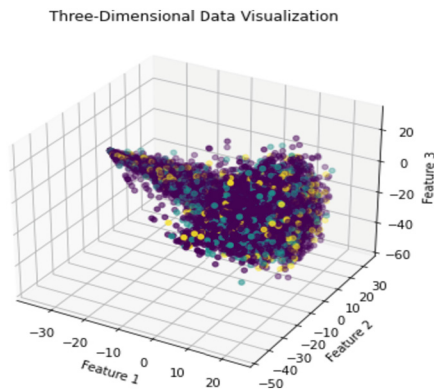

Fig. 1. Data distribution graph.

Figure 1 shows the data distribution graph and Fig. 2 shows a block diagram of our proposed method. The first step represents the data collection and preparation process. In this stage, the COUGHVID dataset was prepared for use in the study. Then, MFCC was extracted from the data. Data more was meaningful and processable in this step. Then, the models were trained using machine SVM, KNN, Logistic Regression, Decision Trees, Random Forest, CNN algorithms. The trained models represented classification models used to classify and categorize the data. These models successfully identified three different classes: COVID19, healthy and symptomatic. Finally, performance analysis was performed to evaluate the model's work.
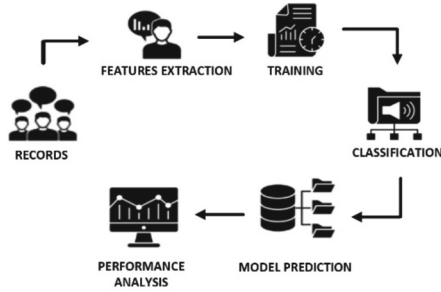
Fig. 2. The flow diagram of the system.

4.1 COVID-19 Detection Using ML Techniques

In this study, COVID-19 disease was detected using 5 different algorithms with machine learning techniques. Pseudocodes of algorithms is given Table 1.

Table 1. Pseudocodes of algorithms.

CNN	SVM

1. Data ◄——the data for the CNN model
2. 'X_train' and 'X_test' to make them compatible with Conv1D input shape.
3. Model ◄—— Create a Sequential
4. Layers ◄—— add layers as follows:
  a. Conv1D layer with 32 filters, kernel size 3, and ReLU activation, X_train.shape[1], 1.
  b. MaxPooling1D layer with pool size 2.
  c. Dropout layer rate 0.25.
  d. Flatten layer ◄—— flatten the data.
  e. Dense layer 64 neurons and ReLU activation.
  f. Dropout layer rate 0.5.
  g. Dense output layer 3 neurons and softmax activation for three classes.
	1. SVM model with 'SVC()' and define a hyperparameter grid for grid search.
	2. Perform grid search using 'GridSearchCV' to find the best hyperparameter combination for the SVM model.
	3. Print the best hyperparameters and the best training accuracy from grid search.
	4. Make predictions on the test set using the SVM model.

Random Forest	**KNN**
1. Create a RandomForestClassifier 'rfc' and fit it on the training data. 2. Make predictions on the test set using the trained RandomForestClassifier 'rfc'. 3. accuracy_score ◄—— Calculate test accuracy. 4. confusion_matrix◄—— Calculate confusion matrix	1. Create a KNeighborsClassifier model 'knn' with 5 neighbors and 'manhattan' distance metric. 2. Fit the 'knn' model on the training data. 3. knn.predict(X_test) ◄—— predictions on the test.
Decision Tree	**Logistic Regression**
1. Create a DecisionTreeClassifier 'dtc' and set up the parameter grid for Grid Search using 'param_grid'. 2. Perform Grid Search using 'GridSearchCV' with 5-fold cross-validation and 'accuracy' scoring. 3. Get the best parameters from the Grid Search and create a new DecisionTreeClassifier 'best_dtc' with these parameters. 4. Fit 'best_dtc' on the training data. 5. Make predictions on the test set using 'best_dtc'.	1. Create a LogisticRegression model 'log' with default parameters and no random state specified. 2. Fit the 'log' model on the training data. 3. log.predict(X_test) ◄—— predictions on the test.

The most accurate successful model of all models was the KNN using Manhattan distance and the accuracy rate was 78.56%. Other KNN models develop with Euclidian and Minkowski distances calculated in 78.36% accuracy. For SVM algorithm, three different kernels were tested on COUGHVID dataset. The best performance for the sigmoid kernel was determined 70.69% accuracy rate and for RBF and LINEAR kernels was computed with 77.04% accuracy rate. The Logistic Regression method also achieved similar results to the other SVM methods with an accuracy rate was 77.04%. The RF model achieved 76.88% accuracy rate. Decision Trees perform the lowest accuracy of all models with an accuracy rate was 60.33%. The results of the ML algorithms were given Table 2.

Table 2. The results of the machine learning algorithms.

Method	Accuracy (%)	Precision (%)	Recall (%)
SVM-RBF	77.04	59.0	77.0
SVM-LINEAR	77.04	59.0	77.0
SVM-Sigmoid	70.69	63.0	71.0
Logistic Regression	77.04	59.0	77.0
Decision Tree	60.33	62.0	60.0
Random Forest	76.88	66.0	77.0
KNN-Minkowski	78.36	64.0	75.0
KNN-Euclidian	78.36	64.0	75.0
KNN-Manhattan	78.56	63.0	75.0

4.2 COVID-19 Detection Using DL Techniques

3 different CNN models with 4 layers were developed in the experimental studies part. Models were evolved with the filter numbers of the layers as "8,16", "16,32", "32,64" respectively. The number of batches for each model is given as 4, 8 and 16 and for each batch number was tested 50 and 100 epochs. In all models, 100 epochs performance was better than 50 epochs performance. The highest performance was 77.04% accuracy rate was obtained in the model with filter numbers is "32,64", batch numbers is 8 and 100 epochs. Training, validation and test accuracies of the models were shown Table 3. The Model 3 accuracy and loss graphs with batch size 8 were shown Fig. 3.

Table 3. Training, validation and test accuracies of the models.

| Model	Number of Baches	Train Accuracy (%)	Validation Accuracy (%)	Test Accuracy (%)	Train Accuracy (%)	Validation Accuracy (%)	Test Accuracy (%)
		50 epoch			**100 epoch**		
M1	4	68.96	69.18	69.18	76.60	76.85	76.85
	8	68.64	69.06	69.06	74.61	74.85	74.85
	16	69.75	68.93	68.93	70.45	70.45	70.78
M2	4	74.48	74.79	74.79	76.28	76.51	76.51
	8	74.61	75.16	75.16	75.48	75.56	75.56
	16	72.73	72.94	72.94	76.11	76.42	76.42
M3	4	76.77	76.91	76.91	76.53	76.64	76.64
	8	**76.50**	**76.61**	**76.61**	**76.61**	**77.04**	**77.04**
	16	76.21	76.08	76.08	76.34	76.57	76.57
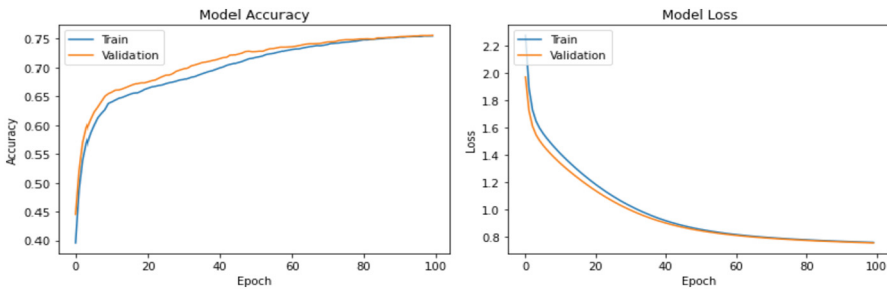
Fig. 3. The Model 3 accuracy and loss graphs with batch size 8.

5 Discussion and Conclusion

This study was conducted to evaluate the usability of both traditional ML and DL techniques for COVID-19 disease detection. In the experimental studies, three different CNN models were developed with four layers and different features were extracted. To evaluate the performance of the model, tests were performed by applying different batch numbers and step numbers. The highest performance was 77.04% accuracy rate in the model filter is "32,64" at batch numbers is 8 and 100 epochs. Also, SVM, RF, Decision Tree, Logistic Regression and KNN was used as machine learning models for classification and regression stage. The highest accuracy rate was recorded as 78.56% using Manhattan distance with KNN. Decision Trees accuracy rate was the lowest performance of 60.33%. Different models and model configurations compare each other and measure their performance with each effect. This study demonstrates the potential of deep artificial intelligence-based methods in Covid-19 diagnosis with the COUGHVID dataset.

References

1. Jahangir, M.A., Muheem, A., Rizvi, M.F.: Coronavirus (COVID-19): history, current knowledge and pipeline medications. J. Pharm. Res. Sci. Technol. [ISSN: 2583–3332], **4**(1), 1–9 (2020)
2. Gudigar, A., et al.: Role of artificial intelligence in COVID-19 detection. Sensors **21**(23), 8045 (2021)
3. Melek, N.: A brief dive into COVID-19 cough sound data sets. In: 2021 Medical Technologies Congress (TIPTEKNO). IEEE (2021)
4. Bozkurt, F., Bayram, E.: Local binary pattern based COVID-19 detection method using chest X-Ray images. In: 2021 29th Signal Processing and Communications Applications Conference (SIU). IEEE (2021)
5. Deshpande, G., Schuller, B.: An overview on audio, signal, speech, & language processing for covid-19 (2020). arXiv preprint arXiv:2005.08579
6. Biçer, M.B.: Meme kanseri görüntülemesinde mikrodalganın yeri." Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi **30**(4), 257–263 (2014)
7. Günay, O., Öztürk, H., Yarar, O.: İyonize Radyasyon ile Çalışan Tıbbi Görüntüleme Cihazlarının Yapısının Proje Temelli Öğrenimi. J. Health Serv. Educ. **3**(1), 20–27 (2019)
8. Despotovic, V., et al.: Detection of COVID-19 from voice, cough and breathing patterns: dataset and preliminary results. Comput. Biol. Med. **138** 104944 (2021)
9. Bagad, P., et al.: Cough against COVID: Evidence of COVID-19 signature in cough sounds. arXiv preprint arXiv:2009.08790 (2020)
10. Aveyard, P., et al.: Association between pre-existing respiratory disease and its treatment, and severe COVID-19: a population cohort study. Lancet Respir. Med. (2021)
11. Sharma, N., et al.: Coswara--A Database of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis. arXiv preprint arXiv:2005.10548 (2020)
12. Dash, T.K., Mishra, S., Panda, G., Satapathy, S.C.: Detection of COVID-19 from speech signal using bio-inspired based cepstral features. Pattern Recogn. **117**, 107999 (2021)
13. Grant, D., McLane, I., West, J.: Rapid and scalable COVID-19 screening using speech, breath, and cough recordings. In: 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), pp. 1–6. IEEE (2021)
14. Coppock, H., Gaskell, A., Tzirakis, P., Baird, A., Jones, L., Schuller, B.: End-to-end convolutional neural network enables COVID-19 detection from breath and cough audio: a pilot study. BMJ Innov. **7**(2) (2021)
15. Chaudhari, G., et al.: Virufy: Global applicability of crowdsourced and clinical datasets for AI detection of COVID-19 from cough (2020). arXiv preprint arXiv:2011.13320
16. Tena, A., Claria, F., Solsona, F.: Automated detection of COVID-19 cough. Biomed. Signal Process. Control **71**, 103175 (2022)
17. Han, J., et al.: Exploring automatic COVID-19 diagnosis via voice and symptoms from crowdsourced data. In: ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8328–8332. IEEE (2021)
18. Alsabek, M.B., Shahin, I., Hassan, A.: Studying the similarity of COVID-19 sounds based on correlation analysis of MFCC. In: 2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI), pp. 1–5 (2020)

19. Celik, G.: CovidCoughNet: A new method based on convolutional neural networks and deep feature extraction using pitch-shifting data augmentation for covid-19 detection from cough, breath, and voice signals. Comput. Biol. Med. 107153 (2023)
20. Ulukaya, S., Sarıca, A.A., Erdem, O., Karaali, A.: MSCCov19Net: multi-branch deep learning model for COVID-19 detection from cough sounds. Med. Biolog. Eng. Comput. 1–11 (2023)
21. Kranthi Kumar, L., Alphonse, P.J.A.: COVID-19 disease diagnosis with light-weight CNN using modified MFCC and enhanced GFCC from human respiratory sounds. Europ. Phys. J. Spec. Top **231**(18), 3329–3346 (2022)
22. Akdeniz, F., Becerikli, Y.: Detection of copy-move forgery in audio signal with mel frequency and delta-mel frequency kepstrum coefficients. In: 2021 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 1–6. IEEE (2021)
23. Chatterjee, S., Kleijn, W.B.: Auditory model-based design and optimization of feature vectors for automatic speech recognition. IEEE Trans. Audio Speech Lang. Process. **19**(6), 1813–1825 (2010)
24. Kurani, A., Doshi, P., Vakharia, A., Shah, M.: A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. Ann. Data Sci. **10**(1), 183–208 (2023)
25. Gao, X., Li, G.: A KNN model based on manhattan distance to identify the SNARE proteins. IEEE Access **8**, 112922–112931 (2020)
26. Raju, C.G., Amudha, V., Sajiv, G.: Comparison of linear regression and logistic regression algorithms for ground water level detection with improved accuracy. In: 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), pp. 1–6. IEEE (2023)
27. De Ville, B.: Decision trees. Wiley Interdisc. Rev. Comput. Stat. **5**(6), 448–455 (2013)
28. Kingsford, C., Salzberg, S.L.: What are decision trees? Nat. Biotechnol. **26**(9), 1011–1013 (2008)
29. Genuer, R., Poggi, J.M., Tuleau-Malot, C., Villa-Vialaneix, N.: Random forests for big data. Big Data Res. **9**, 28–46 (2017)
30. Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S.: Review on convolutional neural networks (CNN) in vegetation remote sensing. ISPRS J. Photogramm. Remote Sens. **173**, 24–49 (2021)
31. Albawi, S., Mohammed, T.A., Al-Zawi, S.: Understanding of a convolutional neural network. In: 2017 International Conference on Engineering and Technology (ICET), pp. 1–6 (2017)
32. Orlandic, L., Teijeiro, T., Atienza, D.: The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms. Sci. Data **8**(1), 156 (2021)