JOURNAL COMPUTER SCIENCES INSTITUTE

JCSI 29 (2023) 399–404
Received: 24 August 2023
Accepted: 28 September 2023

Comparison of Machine Learning Performance on Classification of COVID-19 Cough Sounds Using MFCC Features

Porównanie wydajności uczenia maszynowego w zakresie klasyfikacji odgłosów kaszlu COVID-19 przy użyciu funkcji MFCC

Muhammad Thoriq Hidayat, Mohammad Reza Faisal*, Dwi Kartini, Fatma Indriani, Irwan Budiman, Triando Hamonangan Saragih

Department of Computer Science, Lambung Mangkurat University, Banjarbaru 70714, Indonesia

Abstract

Early detection for COVID-19 has now been widely developed. One of the methods used is cough audio detection. This research aims to classify cough audio. Audio feature extraction is performed using Mel Frequency Cepstral Coefficients (MFCC) to obtain numerical features. Feature classification uses SVM, Random Forest, and Naive Bayes methods. Evaluation is done to find the best classification method. The evaluation results in this study show that SVM Kernel RBF produces the best evaluation value with an AUC value of 0.657715.
Keywords: audio cough; SVM; Random Forest; Naive Bayes

Streszczenie

Wczesne wykrywanie COVID-19 zostało obecnie szeroko opracowane. Jedną z zastosowanych method jest wykrywanie dźwięku kaszlu. Badania te mają na celu klasyfikację dźwięku kaszlu. Ekstrakcję próbek audio wykonano przy użyciu Mel Frequency Cepstral Coefficients (MFCC) w celu uzyskania cech numerycznych. Klasyfikacja cech odbywa się przy użyciu metod SVM, Random Forest i Naive Bayes. Wyniki oceny w tym badaniu pokazują, że SVM Kernel RBF daje najlepszą wartość oceny z wartością AUC wynoszącą 0.657715.
Słowa kluczowe: dźwiękowy kaszel; SVM; losowy las; Naiwny Bayes

*Corresponding author

Email address: : reza.faisal@ulm.ac.id (M. R. Faisal)

1. Introduction

Techniques in detecting Covid-19 in patients have developed from rapid antigens and swab tests to thorax images with the Convolutional Neural Network (CNN) is method [1]. One of the new methods for Covid-19 detection can be done through cough audio. Classification is also applied to cough audio detection as a new method of Covid-19 detection. Cough audio is processed into images and classified with the CNN method [2].

Audio can be converted into other forms, including a spectrum image. Spectrum graphic images are generated from the audio frequency with color to represent the intensity of the signal carried [3]. The graph of the audio signal can be extracted into vectors to be used in classification. The extraction uses Mel Frequency Cepstral Coefficients (MFCC) so that important information related to the audio signal is still carried [4]. MFCC was compared with Zero Crossing Rate (ZCR) and Linear Predictive Coding (LPC) on voice samples, and it found that MFCC had better results [5].

Support vector machine (SVM) research with Linear and Radial Basic Function (RBF) kernels can classify MFCC extraction on Linear kernels with 92.5% accuracy while RBF kernels with 52.6% accuracy [6]. The accuracy value of 83% obtained from the classification results of the SVM method is better than Naive Bayes on unbalanced data. In contrast, the accuracy value of 87% on balanced data obtained from the classification results of the Naive Bayes method is better than SVM.[7]. Based on research [8], music genre detection was performed on music audio data with the random forest method and obtained 80.28% accuracy.

The results obtained from previous research still require further development in the early detection of Covid-19 so that a system is needed that is faster and easier to detect Covid-19. This dataset has never been used for related research before.

The purpose of this research is to determine the performance of machine learning on the classification of covid-19 cough sounds, namely:

1. What is the classification performance of SVM using a combination of parameters and kernels on MFCC extraction?
2. What is the classification performance using a combination of n parameters of the Random Forest estimator?
3. What is the classification performance of the Naïve Bayes model?

2. Literature overview

Random forest classification of music audio genres using music audio data on GitHub, which has 26 features and ten labels with a total of 1000 data, obtained 75% accuracy by determining the number of trees, 72% accuracy obtained from determining three different parameters, 82% accuracy obtained from determining four different parameters [8].

Audio segmentation with MFCC feature extraction uses two types of data: balanced and unbalanced. The best accuracy value for unbalanced data is 0.83 using SVM, and the best value for balanced data is 0.87 using Naïve Bayes [7]. Audio cough detection of coswara and sarcos data amounted to 1079 healthy data and 92 positive data in the Coswara dataset, while 26 negative data and 18 positive data for the Sarcos dataset. The dataset is classified with several algorithms, including LR, SVM, KNN, MLP, CNN, LSTM, and Resnet 50, and then a comparison is made with the best results on Resnet 50 getting an AUC value of 0.96 while SVM is 0.815 [9].

Research by Matin and Valles [3] The accuracy value obtained is 77% on the recognition of autism children's and classified with SVM. Covid-19 detection using the DiCOVA dataset, data in the form of cough audio as many as 1040 records with 75 records of cough sounds of Covid-19 sufferers using SVM and Random Forest classification with Z-score normalization obtained AUC random forest value 82.15 and SVM AUC value 85.05 [10]. Detection of respiratory disorders in COVID patients collected by medical students using MFCC and spectrogram feature extraction obtained CNN accuracy increased from 87.04% to 96.38% [11].

3. Experiment Setting

3.1. Environment and Libraries

The hardware and supporting software specifications used in this research can be seen in Table 1.

Table 1: Specification of the computer

Hardware	
name	Version/description
CPU	Intel(R) Core (TM) i7-9750H
GPU	NVIDIA GeForce GTX 1650
RAM	24 GB DDR4 2666Hz
GPU Driver	NVIDIA 526.98
Software	
OS	Windows 11 pro
package	
matplotlib	3.5.3
scikit-learn	1.0.2
numpy	1.21.6
torch	1.13.0
pandas	1.3.5
librosa	0.10.0.post2

3.2. Classification used

This research will focus on processing cough audio data, extracted into images with Spectrogram and MFCC converts spectrogram images into numeric. Spectrograms images are segmented into frames, each representing the distribution of frequency energy. The frequency domain is transformed into the Mel domain, followed by applying the Discrete Cosine Transform (DCT) to the logarithm of the Mel filter bank, resulting in cepstral coefficients representing the audio signal. These MFCC coefficients serve as features for audio analysis and recognition. formula (1) can be used in convert spectrogram images to numeric with MFCC [12].

$$MFCC(\boldsymbol{n}, m) = \frac{1}{N}\sum_{k=0}^{N-1} \log(\sum_{k=0}^{N-1} X(k,n)e^{-j2\pi km/N})\, H_m(k) \qquad (1)$$

When n is audio frame, m is cepstral coefficient, N is number of frames used in the fast Fourier transform, $X(k,n)$ is FFT value of frame n at frequency k, and $H_m(k)$ is filterbank Mel ke-m at frekuensi-k.

The extracted data will be classified using three classification methods, namely SVM, Random Forest, and Naive Bayes.

In SVM, hyperplane optimization can be done by optimizing parameters. One of the parameters that can be optimized is the parameter C [13]. SVM is a learning system to make predictions in classification cases. SVM can be used formula (2) in linear cases [6].

$$\boldsymbol{K}(x_i, x_j) = x.y \qquad (2)$$

When K is kernel used on SVM, x and y is points in the data that form a vector representing values in the classification.

The use of kernels in SVM can be used for data classification needs that cannot be solved in a linear way. One of these kernels is RBF. the following formula (2) is used in the RBF kernel SVM equation [14].

$$\boldsymbol{K}(x_i, x_j) = exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \qquad (3)$$

When K is kernel used on SVM, x and y is points in the data that form a vector representing values in the classification, and σ is parameters used in the RBF kernel.

Random Forest is a form of decision tree developed as a solution to overfitting in decision trees. Random forest is run by performing the regular random selection. The following formula (3) is used in Random Fores equation [8]. Random Forest is built to minimize bias, and in the construction of each tree, no pruning is done and done randomly [15].

$$Gini\ (S) = 1 - \sum pi2\ k\ i = 1 \qquad (4)$$

When pi is the probability of S belonging to class i

Naive Bayes is one of the classification methods. This method is considered simple and easy to use. In using Naive Bayes, not too much experimental data is needed. Formula (5) is used in Naive bayes equation [7].

$$(C|F_1, \dots, F_n) = \frac{p(C)p(F_1, \dots, F_n|C)}{p(F1, \dots, Fn)} \qquad (5)$$

When $p(C|F_1, \dots, F_n)$ is the posterior probability, $p(C)$ is the prior probability of class C, $p(F_1, \dots, F_n|C)$ is the probability likelihood, and

$p(F1, \ldots, Fn)$ is the prior probability of the instance $(F1, \ldots, Fn)$.

3.3. Used dataset

The dataset used in this research is the COVID-19 Cough Classification data from Alex Wer22ben's Kaggle. The number of cough sound datasets is 1926 audio with two labels, namely 1284 Healthy cough sounds and 642 Covid-19 cough sounds.

Figure 1 and 2 shows the shape of the cough audio data in the dataset used, , with the x-axis displaying the audio frequency value while the y-axis displays the audio time value. Figure 1 is audio with the label healthy, and Figure 2 is audio with the label covid-19.

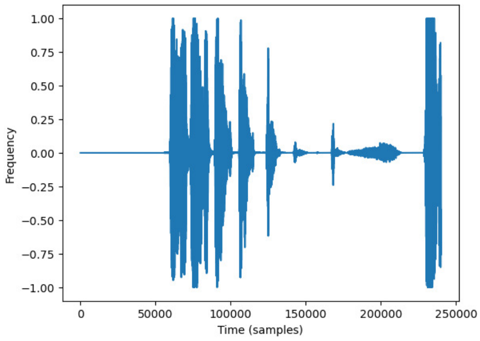

Figure 1: Audio data image healthy label.

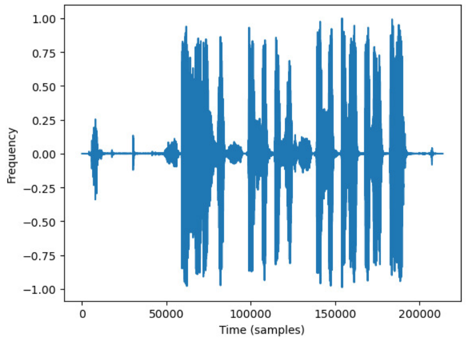

Figure 2: Audio data image covid label.

The audio data is equalized to 5 seconds in duration and then made into a spectrogram image. Figure 3 presents the spectrogram processing results of the audio data with the healthy class.

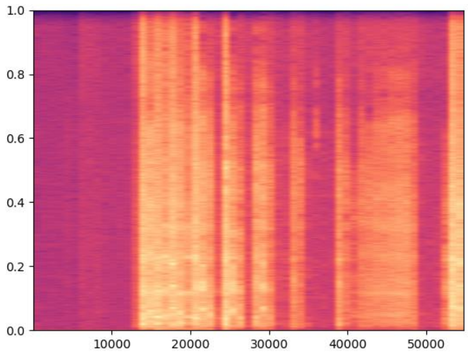

Figure 3: Spectrogram image of labelled healthy audio.

Figure 4 presents the results of processing the Spectrogram of audio data with the covid-19 class. The two-dimensional image produced in the Spectrogram is based on the frequency of the audio.

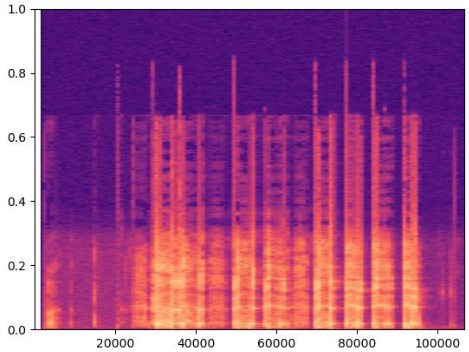

Figure 4: Spectrogram image of covid label.

The colors used in the image represent the altered audio frequency information. The results of the Spectrogram are then extracted into numerical data using MFCC. The resulting data includes 26 features.

3.4. Research Flow

Figure 5 shows the research flow used in this study as follows:

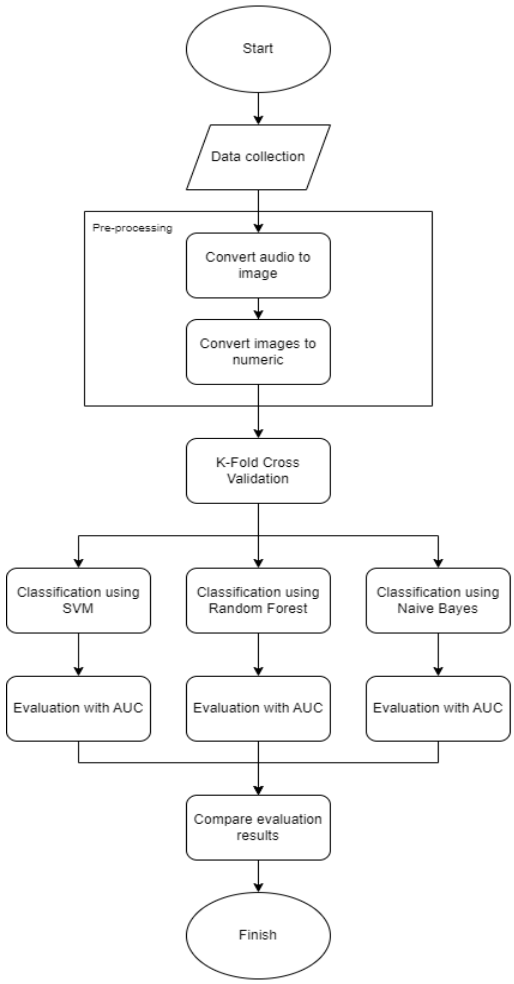

Figure 5: Research flow.

In the initial stage, data will be taken from Kaggle as audio. The data is preprocessed by cutting it with the same duration of five seconds. Once the data has the same duration, the audio data is divided into small segments that represent frequencies and measure their contribution. All segments are assembled into a 2D image with the x-axis representing time, the y-axis representing frequency, and the color or intensity representing the amplitude or power of the frequency component at each time. Darker areas indicate higher frequency energy [16]. The image represents time, the other axis represents frequency, and each point's color indicates the points' amplitude[17]. The image data is extracted again using MFCC to convert the voice data into structured numerical data with 26 extracted features.

Division of training and testing data of MFCC flare feature extraction results using K-fold Cross Validation to minimize overfitting [18]. The K value used is ten, dividing the data into ten partitions with the same number[19]. The classification methods used are SVM, Random Forest, and Naive Bayes.

The classification results are evaluated using the Confusion matrix because the data is not balanced, so the value used as a reference is AUC [20]. Classification evaluation is based on True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) and is calculated according to the performance calculation used [21]. A comparison of AUC value evaluation is carried out to determine the best classification used in the early detection of covid-19 from cough audio.

4. Result

The results of this study were obtained from 4 classification methods and each classification was carried out 10 experiments. The following are the results obtained. Figure 6 presents the AUC results of linear SVM given changes in the value of C from 0.1 to 1.0, and the best results were found in SVM Linear C: 0.9 with an AUC of 0.60. the following confusion matrix results from the best parameters.

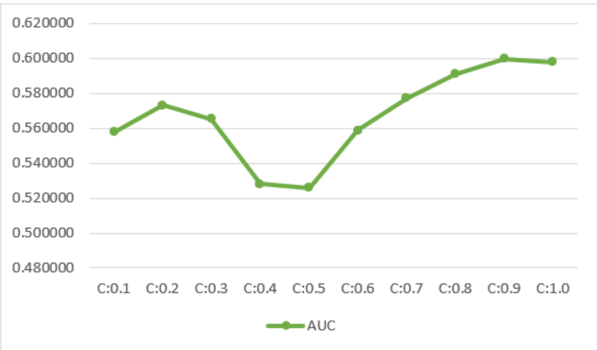

Figure 6: Linear SVM AUC Results.

Figure 7 presents the confusion matrix results of the SVM kernel linear. The following confusion matrix results are obtained from SVM kernel linear classification at the gamma scale parameter and C: 0.9.

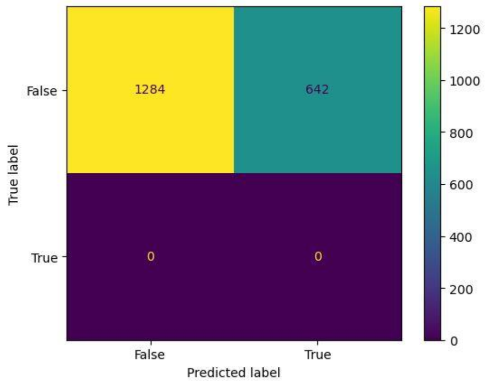

Figure 7: Confusion matrix SVM Linear.

Figure 8 presents the AUC results of SVM RBF given a change in the value of C from 0.1 to 1.0 and found the best results in SVM RBF C: 0.9 with an AUC result of 0.66. the following confusion matrix results from the best parameters.

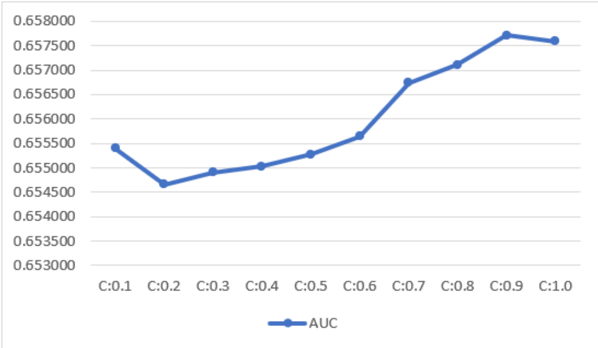

Figure 8: SVM RBF AUC Results.

Figure 9 presents the confusion matrix results of the RBF kernel SVM. The following confusion matrix results are obtained from SVM kernel RBF classification at the gamma scale parameter and C: 0.9.

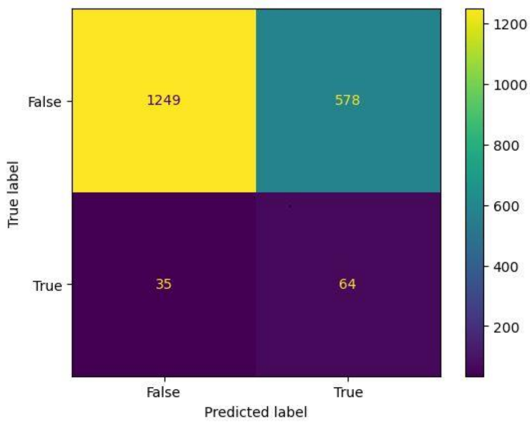

Figure 9: Confusion matrix SVM RBF.

Figure 10 presents the AUC results of Random Forest. The n estimator parameter is a parameter used to determine the number of trees to be made in performing random forest classification. Given changes in the value of the n estimator from 10 to 100, the best results were found in the random forest n estimator 100 with an

AUC result of 0.61. the following confusion matrix results from the best parameters.

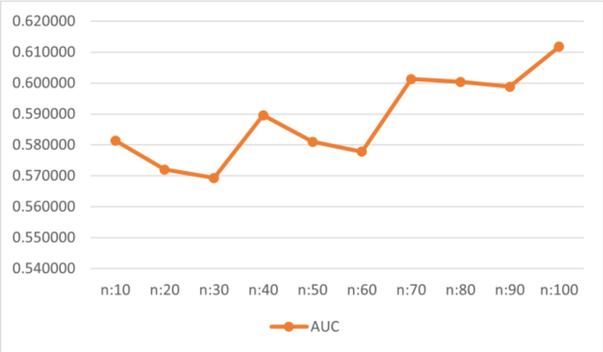

Figure 10: Random Forest AUC Results.

Figure 11 presents the confusion matrix results of Random Forest and those obtained from Random forest classification at parameter n estimator 100.

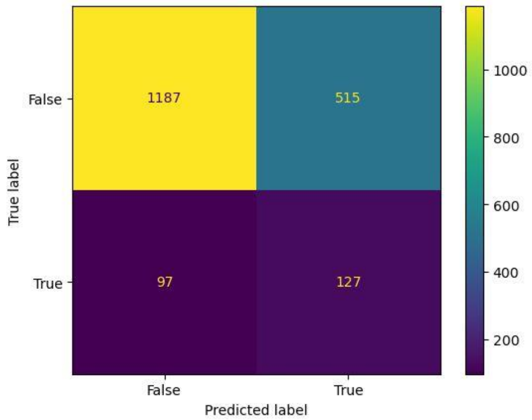

Figure 11: Confusion matrix Random Forest.

Figure 12 below is the confusion matrix result obtained from the Naive Bayes classification, which gets an AUC result of 0.64.

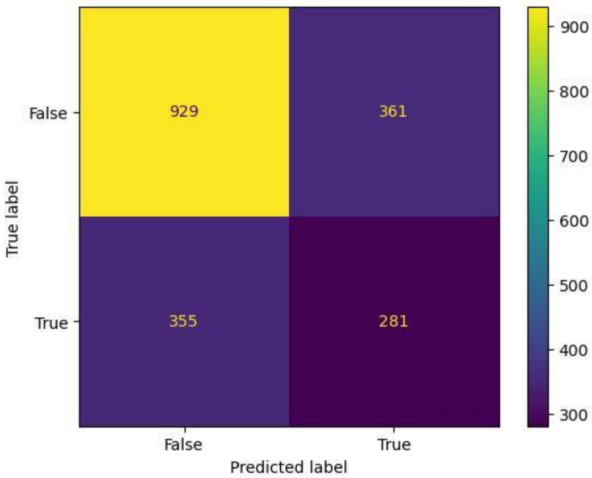

Figure 12: Confusion matrix Naive Bayes.

Table 2 shows the comparison results between the AUC values of each classification method used.

Table 2: Comparison of AUC values

Method	AUC
Support Vector Machine RBF	0.66
Support Vector Machine Linear	0.60
Random Forest	0.61
Naive Bayes	0.64

The order of AUC values in Table 2 is SVM kernel RBF 0.66, Naive Bayes with 0.64, Random Forest with 0.61, and SVM Linear with 0.60. Figure 13 displays a comparison chart of the AUC values of each classification method used.

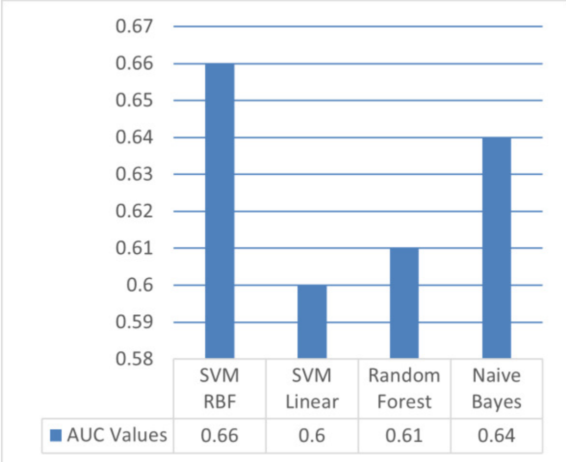

Figure 13: Chart Comparison of AUC values.

5. Conclusions

The conclusion is that using different classifications can produce different values in each AUC value obtained, and changes in parameters can provide changes to the results obtained. In SVM, the changed parameter value increases the value of the results obtained, while in Random Forest, several parameters increase and decrease.

This study found that SVM classification with RBF kernel is considered the best for early detection of covid-19 in cough audio, while for Linear SVM, there is still overfitting based on the confusion matrix results obtained.

For future research, it is recommended to try using deep learning, such as Convolutional Neural Networks (CNN). Further research can also add normalization to the data, using hyperparameters and adding extraction methods. Data can be normalized with min-max and Z-Score. As for hyperparameters, they can be used to get better results. That can be added to the extraction stage in the form of Chroma Constant-Q Transform, Chroma Energy Normalized Statistics, Chroma Variable-Q Transform, Mel Spectrogram, Spectral Contrast, Poly Features, Tonnetz and others.

Acknowledgements

The computation time of the computer system utilized in this study was bestowed by the Data Science Lab of the Computer Science Department, Faculty of Mathematics and Natural Sciences at Lambung Mangkurat

University. This endeavour was sustained by the Program Dosen Wajib Meneliti (PDWM) grant from PNBP Lambung Mangkurat University.

References

[1] S. Hassantabar, M. Ahmadi, A. Sharifi, Diagnosis and detection of infected tissue of COVID-19 patients based on lung X-ray image using convolutional neural network approaches, Chaos Solitons Fractals 140 (2020) 110-170, https://doi.org/10.1016/j.chaos.2020.110170.

[2] V. Bansal, G. Pahwa, N. Kannan, Cough Classification for COVID-19 based on audio mfcc features using Convolutional Neural Networks, in 2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON) IEEE (2020) 604–608, https://doi.org/10.1109/GUCON48875.2020.9231094.

[3] R. Matin, D. Valles, A Speech Emotion Recognition Solution-based on Support Vector Machine for Children with Autism Spectrum Disorder to Help Identify Human Emotions, in 2020 Intermountain Engineering, Technology and Computing (IETC) IEEE (2020) 1–6, https://doi.org/10.1109/IETC47856.2020.9249147.

[4] G. Karaca, Y. Kutlu, Turkish voice commands based chess game using gammatone cepstral coefficients, arXiv preprint arXiv:2101.08441 (2021) https://doi.org/10.48550/arXiv.2101.08441.

[5] N. Chauhan, T. Isshiki, D. Li, Speaker Recognition Using LPC, MFCC, ZCR Features with ANN and SVM Classifier for Large Input Database, in 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), IEEE (2019) 130–133, https://doi.org/10.1109/CCOMS.2019.8821751.

[6] S. R. Chaudhary, S. N. Kakarwal, J. V. Bagade, Feature selection and classification of indian musical string instruments using SVM, Indian Journal of Computer Science and Engineering 12(4) (2021) 859–867, https://doi.org/10.21817/indjcse/2021/v12i4/211204142.

[7] L. O. Iheme, Ş. Ozan, Multiclass digital audio segmentation with MFCC features using naive Bayes and SVM classifiers, in 2019 Innovations in Intelligent Systems and Applications Conference (ASYU) (2019) 1–5, https://doi.org/10.1109/ASYU48272.2019.8946441.

[8] N. Ndou, R. Ajoodha, A. Jadhav, Music genre classification: A review of deep-learning and traditional machine-learning approaches, in 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS) (2021) 1–6, https://doi.org/10.1109/IEMTRONICS52119.2021.9422487.

[9] M. Pahar, M. Klopper, R. Warren, T. Niesler, COVID-19 cough classification using machine learning and global smartphone recordings, Comput Biol. Med. 135 (2021) https://doi.org/10.1016/j.compbiomed.2021.104572.

[10] I. Södergren, M. P. Nodeh, P. C. Chhipa, K. Nikolaidou, G. Kovács, Detecting COVID-19 from Audio Recording of Coughs Using Random Forests and Support Vector Machines, in Interspeech 2021, ISCA (2021) 916–920, https://doi.org/10.21437/Interspeech.2021-2191.

[11] M. M. Gauy, M. Finger, Audio MFCC-gram Transformers for respiratory insufficiency detection in COVID-19, arXiv preprint arXiv:2210.14085 (2022) https://doi.org/10.48550/arXiv.2210.14085.

[12] M. B. Alsabek, I. Shahin, A. Hassan, Studying the Similarity of COVID-19 Sounds based on Correlation Analysis of MFCC, in 2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI), IEEE (2020) 1–5, https://doi.org/10.1109/CCCI49893.2020.9256700.

[13] A. S. Elkorany, M. Marey, K. M. Almustafa, Z. F. Elsharkawy, Breast Cancer Diagnosis Using Support Vector Machines Optimized by Whale Optimization and Dragonfly Algorithms, IEEE Access 10 (2022) 69688–69699, https://doi.org/10.1109/ACCESS.2022.3186021.

[14] A. Razaque, M. Ben Haj Frej, M. Almi'ani, M. Alotaibi, B. Alotaibi, Improved support vector machine enabled radial basis function and linear variants for remote sensing image classification, Sensors 21(13) (2021) 4431, doi: https://doi.org/10.3390/s21134431.

[15] E. Kubera, A. Wieczorkowska, A. Kuranc, and T. Słowik, Discovering Speed Changes of Vehicles from Audio Data, Sensors 19(14) (2019) 3067, https://doi.org/10.3390/s19143067.

[16] K. Palanisamy, D. Singhania, and A. Yao, Rethinking CNN models for audio classification, arXiv preprint arXiv:2007.11154, (2020) https://doi.org/10.48550/arXiv.2007.11154.

[17] Y. Zeng, H. Mao, D. Peng, Z. Yi, Spectrogram based multi-task audio classification, Multimed Tools Appl, 78(3) (2019) 3705–3722, https://doi.org/10.1007/s11042-017-5539-3.

[18] N. Fazakis, V. G. Kanas, C. K. Aridas, S. Karlos, S. Kotsiantis, Combination of active learning and semi-supervised learning under a self-training scheme, Entropy 21(10) (2019) 988, https://doi.org/10.3390/e21100988.

[19] D. Berrar, Cross-Validation, in Encyclopedia of Bioinformatics and Computational Biology, Tokyo: Elsevier (2019) 542–545, https://doi.org/10.1016/B978-0-12-809633-8.20349-X.

[20] V. Maeda-Gutiérrez *et al.*, Comparison of convolutional neural network architectures for classification of tomato plant diseases, Applied Sciences 10(4) (2020) 1245 https://doi.org/10.3390/app10041245.

[21] J. R. Maria Navin, R. Pankaja, Performance analysis of text classification algorithms using confusion matrix, International Journal of Engineering and Technical Research (IJETR) 6(4) (2016) 75–78.