Computer Science Journal of Moldova, vol.**32**, no.2(95), 2024

Machine Learning and Sound Processing in Vocal Disease Detection

Mihai-Andrei Costandache

Abstract

We present in this paper some of the existing machine learning and sound processing techniques involved in the medical process and show how they can be applied in the context of a vocal disease detection task. The machine learning techniques are the usual, ranging from decision trees to neural networks, and when it comes to the sound processing approaches some scientists probably are familiar with them from the speech processing or music analysis areas. However, the techniques are adapted to the particularities of the medical field – data collection, verification, etc. Through the concrete example of the vocal disease detection task we worked on, we made some interesting observations, both on the classification ability of the model and on the influence of the data.

Keywords: machine learning, sound processing, medicine.
MSC 2020: 68T07.

1 Introduction

The technological progress provided interesting changes in the medical field. Automated systems are not new for medical experts; what we consider to be different is the introduction of machine learning in tasks that so far have been done only by human experts. For example, a machine learning system can analyze audio samples and provide a verdict. Interestingly, we have not reached yet the stage where the systems can replace human expertise fully, and we consider that it would also not be appropriate. A mixed approach where the doctors and the computer

doi:10.56415/csjm.v32.12

systems work together is more indicated, as the doctors can focus on the special scenarios or the patient particularities.

Section 2 presents machine learning and sound processing in general and explains briefly the main techniques, without being limited to the ones found in the cited articles. Section 3 shows some of the machine learning and sound processing solutions we found for identifying diseases. Most of the solutions are presented by the following template: machine learning algorithm – sound features. In this way, we get a summarised yet interesting view. The methods are not specific to a particular medical condition, but it was natural to classify them by the disease type. Section 4 approaches the solutions for treatment. A significant difference from the diagnosis part is that the solutions use other techniques instead of sound processing (audio samples are more appropriate in the diagnosis part). Section 5 presents some experiments on a vocal disease data collection. We explored how to build a classifier model and how the limited number of instances of a medical condition affects the machine learning model. Section 6 summarises the paper.

2 Overview of Machine Learning and Sound Processing

Artificial intelligence is a term that is difficult to define. [1] – [2] described artificial intelligence as the imitation by machines of the intelligence in humans. Machine learning is a field of artificial intelligence, representing the ability of machines to learn without being explicitly programmed, according to [3] and [4]. From [3] and several other cited papers, we present some of the most popular methods of supervised learning, consisting in obtaining a mapping from example input-output pairs.

- *Decision Tree* – tree graph representation of choices and results;
- *Support Vector Machine (SVM)* [5] – method that searches for the hyperplane with the maximum distance to the nearest instance of each class;

- *K-Nearest Neighbors (k-NN)* – classification based on proximity to each instance;
- *Naive Bayes* [6] – classification based on Bayes Theorem.

There are also the concepts of reinforcement learning (maximizing a cumulative reward), generative models, ensemble learning, boosting, and bagging. Some of the terms refer to the aggregation of multiple algorithms, not a specific algorithm. The neural networks [7] – [8] are a technique that we approach separately. Basically, from multiple inputs and weights, an activation function retrieves an output. A special type is the *Convolutional Neural Network (CNN)* [8], suited for image classification and natural language processing. The sound data are often transformed into a visual representation, image processing is relevant in this context.

Signal processing [9] consists in mapping or transforming signals, and it can be:

- *Analog* – continuous functions;
- *Digital* – discrete functions (the approach in the computer processing context).

Sampling [9] is the process of obtaining a digital signal from an analog one. Sample rate [10] – [11], measured in Hertz (Hz), signifies the number of samples per second extracted from a continuous signal. In Hz it is also measured the frequency (or pitch) [12], representing intuitively how high or low the sound is (e.g., the human hearing range is 20-20000 Hz). Nyquist-Shannon Theorem [13] – [14] expresses a relationship between the highest frequency and the sample rate, it states that

> *If a function $x(t)$ contains no frequencies higher than B hertz, then it can be completely determined from its ordinates at a sequence of points spaced less than $1/(2B)$ seconds apart.*

Fourier Transform [9] – [10] changes the domain of a continuous function from time to frequency, Discrete Fourier Transform (DFT) [9] is its discrete equivalent, Short Time Fourier Transform [10], [15] is the transformation on segments, and Fast Fourier Transform [16] is an algorithm to compute DFT or its inverse.

Mel scale [17] – [19] is a subjective scale of pitches, perceived by listeners as equally distanced. There are multiple formulas for the Mel scale, one of them [18] is:

$$m = 2595lg(1 + f/700), \tag{1}$$

m is the number of *mels* and f is the number of *hertz*.

Some of the features [10], [20] – [21] that can be extracted from sound are:

- *Zero-crossing rate (ZCR)* [22] – a measure of how many times the amplitude of the signal is a value of zero (in an interval / frame);
- *Spectral centroid* [23] – a measure of the amplitude at the spectrum center of the signal;
- *Mel-frequency cepstral coefficients (MFCC)* [19] – features that describe the shape of the spectrum.

Some of the image representations [20] – [21] of sound are:

- *Waveform* – representation that indicates the frequency and amplitude of the sound;
- *Chromagram* – representation that expresses how much energy each frequency corresponding to a musical note has;
- *Spectrogram* [24] – representation with time on the X-axis, frequency on the Y-axis, and amplitude coded by color.

3 Diagnosis Solutions

3.1 Respiratory Diseases

The sound indicators from which a respiratory medical condition can be assumed are breathing, voice, and coughing.

In [25], the authors used a CNN and a bidirectional long short-term memory (LSTM) combined model to identify diseases from lung sound: asthma, pneumonia, bronchiectasis (BRON), chronic obstructive pulmonary disease (COPD), and heart failure (HF). The preprocessing steps included wavelet smoothing, displacement removal, and normalisation.

Paper [26] approached tuberculosis detection from coughing sound. The extracted features were MFCC, log-filterbank energies, ZCR, and kurtosis. Five classifiers were applied separately, with multiple extracted features configurations: logistic regression, SVM, k-NN, multilayer perceptrons, and CNN. [27] used a CNN to identify COPD from lung sound. They applied data augmentation techniques (loudness, mask, shift, speed edits) and extracted mel-spectrograms, MFCC, and chromagrams.

In [28], the authors created a two-stage system. It analyzed whether the patient had a respiratory medical condition, then found the condition (two diseases were recognized, and the rest of the diseases were labeled as other). There was not a particular machine learning algorithm used, experiments on multiple techniques were run. The feature configuration was determined automatically (one of the criteria was based on the Gini Index).

In [29], the authors built a solution for the diagnosis of some respiratory diseases with CNNs. Mel-spectrograms were extracted from coughing sound samples. [30] tried multiple techniques, comparing SVM, k-NN, and Gaussian Bayes in the diagnosis of respiratory diseases. The features were MFCC from lung sound and sometimes text data (patient information). [31] used transfer learning and a CNN to classify breathing sound from mel-spectrograms.

3.2 Heart Diseases

The indicator is the sound of the heart cycles. [32] remarked that recording heart sounds is challenging due to the complexity and diversity in the signal manifestation and to the weakness of heart sounds, sometimes covered by noise.

In [33], the authors proposed an Internet of Health Things (IoHT) system for diagnosis. They preprocessed the data by resampling, normalising, and filtering, then performed classification using an autoencoder neural network.

In [34], the authors extracted features with MFCC and DWT techniques, sometimes combined, and for classification, they tried SVM, deep neural networks, and centroid displacement-based k-NN. [35] used a deep neural network, the extracted features were chromagram, spectral centroid, spectral bandwidth, ZCR, and MFCC.

In [36], a technique that skips feature extraction and segmentation in heart sound classification was created, using a Transformer and CNN on the input signals.

In [37], transfer learning and CNNs (the authors compared some architectures) were used to classify heart sound. Mel-spectrograms were the extracted features. [38] also used transfer learning and CNNs, but with repetition-based spectrograms.

3.3 Sleep Diseases

The indicator is the breathing sound. [39] reviewed the detection of sleep apnea – medical condition consisting in breathing difficulties during sleep. The authors remarked that full night polysomnography (PSG), physiological signals collection made in a laboratory, is uncomfortable, expensive, and time consuming. Some works use adapted, less complicated measurements.

In [40], the authors proposed techniques for assigning severity groups in terms of sleep disordered breathing to patients. The authors tried simple logistics, SVM, and a deep neural network, on multiple extracted features. [41] used a CNN to detect sleep apnea from the SpO2 signal. They compared their approach to linear discriminant analysis (LDA), SVM, bagging representation tree, and an artificial neural network, methods existing in the literature. [42] used SVM to diagnose sleep apnea into several categories from sound, on multiple recording types: single-channel SpO2, single channel airflow, and dual-channel of simultaneous SpO2 and airflow.

In [43], the authors created a snore detection system with a CNN that processes the sound signals and one for the images obtained from a visibility graph.

In [44], the tasks of making the sleep/wake distinction and diagnosing sleep apnea were approached, there were used deep neural networks and spectrograms. [45] detected snore with a CNN and recurrent neural network layer combination.

3.4 Speech/Mental Diseases

The indicator is the voice and we approached the diseases in the same category because they are sometimes connected. The medical conditions can be related to the vocal fold or psychological.

In [46], the authors tried to diagnose voice disorders (e.g., neoplas) using SVM, random forest, k-NN, gradient boosting, and ensemble learning that combined the approaches, from MFCC. [47] proposed using neural networks for the estimation of vocal fold physiology. [48] evaluated several algorithms (e.g., bidirectional LSTM) for pathological voice problem diagnosis from MFCC.

In [49], the authors used an LSTM and an autoencoder, processing spectrograms, for pathological voice problem detection (e.g, depression). [50] and [51] approached the detection of Parkinson's disease. Fig. 1 shows some signals.

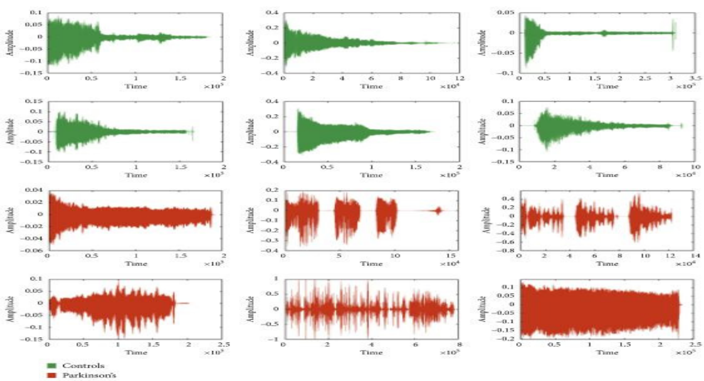

Figure 1. Amplitudes of controls and Parkinson's subjects [50]

Paper [52] approached depression detection.

4 Treatment Solutions

Paper [53] described precision medicine, an approach that focuses on understanding and treating diseases by using data from an individual to make customized decisions. The data consists of medical diagnoses, clinical phenotype, biologic investigations, environment, lifestyle, etc. Fields such as genetics and pharmacology are involved.

In [54], the authors used machine learning to study if corticosteroids and remdesivir work against COVID-19. [55] proposed collaborative filtering and clustering techniques for drug recommendation to diabetes patients. [56] used an artificial neural network for drug recommendation in the context of infectious diseases.

In [57], the authors approached the intraoperative hypotension problem.

5 Vocal Disease Detection Task

The primary objective in this task was to make a speech file classifier, that recognized if the speaker was healthy or what disease (Laryngocele or Vox senilis) he/she had. Theoretically a model can be extended / retrained and have minor adjustments, if more medical conditions are available in new data, so starting with the two diseases was reasonable. The secondary objective was to analyse the effects of data that were missing a class balance.

We made several experiments on a sound recordings database from [58], challenging from the perspective of class representation (number of examples), allowing us to get interesting insights and observe certain patterns. The origin of the data is [59], that is a very interesting and useful resource for machine learning and medical work. On the website in [59] the user can request files based on age, sex, normal or pathological characteristics, diagnosis of the speakers in recordings, or the specific ids of sessions or speakers. The user can preview the recordings after he/she queries the data, and find out information about the session id, patient id, diagnosis (multiple are permitted), etc. A patient can have more than one session. The files can be exported through an archive, where each filename reflects the id, what it contains, and whether there is from an EGG. More information about the files seems to missing in the archive but by observation we consider the id to be that of the session. It would have been interesting to have the website include the patient id too, from our knowledge it is not possible (or at least it is not possible by default, or it was not the case in the customised database). The user can manually compare the previewed files and afterwards associate ids from the website, but in a machine learning context of large data collections, it is not convenient. Regarding the EGG, it is not explained on the website, but from our research the files must come from an instrument called electroglottograph [60], generating a high frequency electric current. Two electrodes put externally around the thyroid area are used. Besides the distinction between regular and EGG files, a file can contain particular sounds, pronounced with low / normal / high pitch or in sliding production mode, or a whole phrase.

The data consists of some classes representing normal and pathological characteristics, the regular and EGG files are equally represented, there are 14 combinations of input in a session. Table 1 shows how many examples are there.

Table 1. Vocal Disease Detection Data

	Total	Regular/EGG each	Sessions
Laryngocele	84	42	3
Normal	560	280	20
Vox senilis	392	196	14

From a quick verification, there do not seem to be sessions appearing with more than one class and from the website sessions with both Laryngocele and Vox senilis were not found, although in general a session can have multiple assigned diagnoses.

We interpreted the data in the following way. Firstly, we considered that regular and EGG files should be separated. Listening to files of both types, it was clear there were different characteristics; therefore, it was logical to have a separate machine learning model for each type. In a real context, the medic must send recordings to the appropriate model, considering how he/she obtained the files. However, combining the files is reasonable even if we did not choose this approach. We compared the regular and EGG files.

Secondly, we explored if splitting the data by the session id (a session has all the files in train or in test, for greater test independency) or placing no restriction (splitting at the file level) makes a difference. Another data splitting idea was to do for the patients what we did for the sessions. However, besides the fact that the information in the exported files that would have allowed us to know the patient was missing, we considered such a data split was also not necessary. Because of the session differences (e.g, patient evolution), the similarities between files of the same patient but different session should be fewer than the similarities between files in the same session.

In the splitting process, selecting by id the samples to include in the test collection was with a different granularity than selecting by file. Also, to represent all classes in the test, if there was no Laryngocele session in the data, one was added by default. As there can be seen when the parameters are presented, the split was approximately 0.25. There were 37 sessions, splitting into 4 and applying ceil we had 10 test sessions. Each session had 14 files. In order not to have differences in numbers, selecting by file also required 140 test files. Based on the no missing class requirement, Laryngocele had 14 test files (we avoided more) in the split by id version, but in the split by file version it should have had on average 42 / 4 = 10.5 files. Therefore, splitting by id might lead to more Laryngocele files in the test part, and if this class is difficult to predict, it affects the accuracy.

We classified the files with a CNN (TensorFlow [61], Keras [62]) processing mel-spectrograms (Librosa [20], [21]). Sound/image processing functions were applied (conversion to decibels, image resize). The files were not segmented, they were generally short. Table 2 contains several parameters, Table 3 shows the model.

The model was compiled with the optimizer Adam and the loss function sparse_categorical_crossentropy.

For reproducibility, we used seeds in our experiments. Randomness

Table 2. Configuration

Key	Value
test split	0.25
n_fft	2048
n_mels	128
hop_length	512
image height/width	128
epochs	20
batch size	32
training rate	0.001

Table 3. Architecture

Layer
Input
Rescaling(1./255)
Conv2D(32, (3, 3), activation='relu')
MaxPooling2D((2, 2))
Conv2D(64, (3, 3), activation='relu')
MaxPooling2D((2, 2))
Flatten
Dense(128, activation='relu')
Dense(8, activation='relu')
Dense(3, activation='softmax')

was involved the most in data partitioning and model training.

Analyzing the seeds leading to an appropriate data representation and results (8, 10, etc), we made the following observations:

- There was a good accuracy, on the training files approximately 87% – 98%, on the test files 67% – 85%;
- There was an advantage of regular over EGG files but not in all cases;
- Similar advantage for split by file over by id, possible reasons were allowing connections between session files (some in train, some in test parts) and less Laryngocele examples in some cases.

The results show the good capacity of the model to classify files.

However, some seeds led to unpredictable results as in some situations:

- The model was getting stuck at a small accuracy – even in training, it converged to approximately few percents over 50%, or there was a surprisingly low number of examples in test data for a class (the problems can be connected);
- It was very difficult to make a comparison of regular / EGG files, split by file / id.

The great changes due to the randomness factor (especially the data split) reflect the problems of the data classes. Overall it would be interesting to have tests on a collection with an appropriate data representation, but we consider our work to show data patterns and represent an example of how machine learning and sound processing work against diseases.

6 Conclusion

This paper presented some of the existing machine learning and sound processing techniques in the medical field and our own work on vocal disease detection. The techniques are diverse, and we noticed that the most popular are neural networks and MFCC. We consider that the neural networks tend to represent the usual choice of researchers due to accessibility reasons, both theoretical – they are intuitive to scientists with less machine learning experience too, and practical – with minimal data preprocessing, a working model is obtained. The argument for MFCC is that multiple sound characteristics are represented. In our work, we used mel-spectrograms. The methods in the presented articles and our solution lead to the conclusion that machine learning and sound processing, used in multiple work areas, are also appropriate in medicine. However, the model is as good as the data in some situations; we studied the effects of having a challenging dataset (the model is still successful in general). We need to make the following warning: the medical process has more consequences than other tasks, therefore the involvement of human experts is recommended.

References

[1] H. Hassani, E. S. Silva, S. Unger, M. T. Mazinani, and S. MacFeely, "Artificial Intelligence (AI) or Intelligence Augmentation (IA): What Is the Future?," *AI*, vol. 1, no. 2, April 2020.

[2] H. Sheikh, C. Prins, and E. Schrijvers, "Artificial Intelligence: Definition and Background," in *Mission AI*, Springer, 2023.

[3] B. Mahesh, "Machine Learning Algorithms - A Review," *International Journal of Science and Research*, vol. 9, no. 1, January 2020.

[4] *Machine Learning. Algorithms, Models and Applications* (Artificial Intelligence, vol. 7), IntechOpen, 2021.

[5] T. Evgeniou and M. Pontil, "Support Vector Machines: Theory and Applications," in *Machine Learning and Its Applications*, Advanced Lectures, 2001.

[6] freeCodeCamp, "Bayes' Rule Explained For Beginners," Available: https://www.freecodecamp.org/news/bayes-rule-explained/. Accessed May 28, 2023.

[7] R. Dastres and M. Soori, "Artificial Neural Network Systems," *International Journal of Imaging and Robotics*, vol. 21, no. 2, March 2021.

[8] A. Ghosh, A. Sufian, F. Sultana, A. Chakrabarti, and D. De, "Fundamental Concepts of Convolutional Neural Network," in *Recent Trends and Advances in Artificial Intelligence and Internet of Things*, 2020.

[9] I. Apolinario and P. S. R. Diniz, "Chapter 1. Introduction to Signal Processing Theory," in *Signal Processing Theory and Machine Learning*, 2014.

[10] M.-A. Costandache, M.-A. Cioată, and A. Iftene, "Automated Heart Murmur Detection using Sound Processing Techniques," in *Proceedings of the 27th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems*, September 6-8, 2023, Athens, Greece.

[11] Adobe, "Sample rates and audio sampling: a guide for beginners," Available: https://www.adobe.com/uk/creativecloud /video/discover/audio-sampling.html. Accessed May 28, 2023.

[12] Widex, "The Human Hearing Range - What Can You Hear?," Available: https://www.widex.com/en/blog/global /human-hearing-range-what-can-you-hear/. Accessed May 28, 2023.

[13] C. E. Shannon, "Communication in the presence of noise," in *Proceedings of the Institute of Radio Engineers*, 1949.

[14] H. Nyquist, "Certain topics in telegraph transmission theory," *Trans. AIEE*, 1928.

[15] "Short Time Fourier Transform lecture," Available: https://course.ece.cmu.edu/˜ece491/lectures/ L25/STFT_Notes_ADSP.pdf. Accessed May 28, 2023.

[16] "Fast Fourier Transform," Available: https://jakevdp.github.io/ blog/2013/08/28/understanding-the-fft/. Accessed May 28, 2023.

[17] S. S. Stevens *et al.*, “A scale for the measurement of the psychological magnitude pitch,” *Journal of the Acoustical Society of America*, vol. 8, no. 3, 1937.

[18] D. O’Shaugnessy, “Speech communication: human and machine,” 1987.

[19] “Music Information Retrieval,” Available: https://musicinformationretrieval.com/index.html. Accessed May 28, 2023.

[20] “librosa features,” Available: https://librosa.org/doc/latest/feature.html. Accessed May 28, 2023.

[21] B. McFee *et al.*, “librosa: Audio and music signal analysis in python,” in *Proceedings of the 14th python in science conference*, 2015.

[22] D. S. Shete and S. B. Patil, “Zero crossing rate and Energy of the Speech Signal of Devanagari Script,” *IOSR Journal of VLSI and Signal Processing*, vol. 4, 2014.

[23] V. K. Harpale and V. K. Bairagi, “Seizure detection methods and analysis,” in *Brain Seizure Detection and Classification Using EEG Signals*, Academic Press, 2022.

[24] PNSN, “What is a spectrogram?,” Available: https://pnsn.org/spectrograms/what-is-a-spectrogram. Accessed May 28, 2023.

[25] M. Fraiwan, L. Fraiwan, M. Alkhodari, and O. Hassanin, “Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory,” *Journal of Ambient Intelligence and Humanized Computing*, April 2021.

[26] M. Pahar, M. Klopper, B. Reeve, G. Theron, R. Warren, and T. Niesler, “Automatic Cough Classification for Tuberculosis Screening in a Real-World Environment,” *Physiological Measurement*, vol. 42, no. 10, November 2021.

[27] A. Srivastava, S. Jain, R. Miranda, S. Patil, S. Pandya, and K. Kotecha, "Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease," *PeerJ Computer Science*, February 2021.

[28] O. Karaarslan, K. D. Belcastro, and O. Ergen, "Respiratory sound-base disease classification and characterization with deep/machine learning techniques," *Biomedical Signal Processing and Control*, vol. 87, January 2024.

[29] C. Bales *et al.*, "Can Machine Learning Be Used to Recognize and Diagnose Coughs?," 2020.

[30] M. Aykanat, O. Kilic, B. Kurt, and S. Saryal, "Lung disease classification using machine learning algorithms," *International Journal of Applied Mathematics Electronics and Computers*, December 2020.

[31] Y. Kim *et al.*, "Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning," *Scientific Reports*, August 2021.

[32] S. Li *et al.*, "A Review of Computer-Aided Heart Sound Detection Techniques," *BioMed Research International*, 2020.

[33] O. Deperlioglu, U. Kose, D. Gupta, A. Khanna, and A. K. Sangaiah, "Diagnosis of heart diseases by a secure Internet of Health Things system based on Autoencoder Deep Neural Network," *Computer Communications*, vol. 162, October 2020.

[34] Yaseen *et al.*, "Classification of Heart Sound Signal Using Multiple Features," *Applied Sciences Journal*, vol. 8, no. 12, 2018.

[35] L. Brunese, F. Martinelli, F. Mercaldo, and A. Santone, "Deep learning for heart disease detection through cardiac sounds," *Procedia Computer Science*, 2020.

[36] J. Cheng and K. Sun, "Heart Sound Classification Network Based on Convolution and Transformer," *Sensors*, October 2023.

[37] U. Mukherjee and S. Pancholi, "Heartbeat Sound Classification with Visual Domain Deep Neural Networks," 2021.

[38] V. Arora *et al.*, "Transfer Learning Model to Indicate Heart Health Status Using Phonocardiogram," *Computers, Materials & Continua*, vol. 69, no. 3, August 2021.

[39] S. S. Mostafa, F. Mendonca, A. G. Ravelo-Garcia, and F. A. Morgado-Dias, "Systematic Review of Detecting Sleep Apnea Using Deep Learning," *Sensors*, November 2019.

[40] T. Kim, J.-W. Kim, and K. Lee, "Detection of sleep disordered breathing severity using acoustic biomarker and machine learning techniques," *BioMedical Engineering OnLine*, February 2018.

[41] H. T. Chaw, S. Kamolphiwong, and K. Wongsritrang, "Sleep apnea detection using deep learning," 2019.

[42] D. Alvarez *et al.*, "A machine learning-based test for adult sleep apnoea screening at home using oximetry and airflow," *Scientific Reports*, 2020.

[43] R. Li, W. Li, K. Yue, R. Zhang, and Y. Li, "Automatic snoring detection using a hybrid 1D–2D convolutional neural network," *Scientific Reports*, 2023.

[44] H. Nakano, T. Furukawa, and T. Tanigawa, "Tracheal Sound Analysis Using a Deep Neural Network to Detect Sleep Apnea," *Journal of Clinical Sleep Medicine*, 2019.

[45] J. Xie *et al.*, "Audio-based snore detection using deep neural networks," *Computer Methods and Programs in Biomedicine*, March 2021.

[46] M. Pham, J. Lin, and Y. Zhang, "Diagnosing Voice Disorder with Machine Learning," in *Proceedings of the IEEE International Conference on Big Data*, 2018.

[47] Z. Zhang, "Estimation of vocal fold physiology from voice acoustics using machine learning," *Journal of the Acoustical Society of America*, March 2020.

[48] S.-S. Wang, C.-T. Wang, C.-C. Lai, Y. Tsao, and S.-H. Fang, "Continuous Speech for Improved Learning Pathological Voice Disorders," 2022.

[49] D. Sztaho, K. Gabor, and T. M. Gabriel, "Deep Learning Solution for Pathological Voice Detection using LSTM-based Autoencoder Hybrid with Multi-Task Learning," in *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies*, 2021.

[50] M. Pramanik, R. Pradhan, P. Nandy, S. M. Qaisar, and A. K. Bhoi, "Assessment of Acoustic Features and Machine Learning for Parkinson's Detection," *Journal of Healthcare Engineering*, 2021.

[51] M. Hoq, M. N. Uddin, and S.-B. Park, "Vocal Feature Extraction-Based Artificial Intelligent Model for Parkinson's Disease Detection," Diagnostics, 2021.

[52] H. Negi, T. Bhola, M. S. Pillai, and D. Kumar, "A Novel Approach for Depression Detection Using Audio Sentiment Analysis," in *Proceedings of 4th International Conference on Computers & Management*, 2018.

[53] S. J. MacEachern and N. D. Forkert, "Machine learning for precision medicine," *Genome*, 2021.

[54] C. Lam *et al.*, "Machine Learning as a Precision-Medicine Approach to Prescribing COVID-19 Pharmacotherapy with Remdesivir or Corticosteroids," *Clinical Therapeutics*, May 2021.

[55] L. F. G. Morales, P. Valdiviezo-Diaz, R. Reategui, and L. Barba-Guaman, "Drug Recommendation System for Diabetes Using a Collaborative Filtering and Clustering Approach: Development and Performance Evaluation," *Journal of Medical Interest Research*, vol. 24, no. 7, 2022.

[56] U. Bhimavarapu, N. Chintalapudi, and G. Battineni, "A Fair and Safe Usage Drug Recommendation System in Medical Emergencies by a Stacked ANN," *Algorithms*, 2022.

[57] M. Wijnberge *et al.*, "The use of a machine-learning algorithm that predicts hypotension during surgery in combination with personalized treatment guidance: Study protocol for a randomized clinical trial," *Trials*, 2019.

[58] S. Chakraborty, "Patient Health Detection using Vocal Audio," Available: https://www.kaggle.com/datasets/subhajournal/patient-health-detection-using-vocal-audio. Accessed June 14, 2024.

[59] M. Putzer and W. J. Barry, "Saarbrucken voice database," Available: https://stimmdb.coli.uni-saarland.de/. Accessed June 14, 2024.

[60] UNED Voice Lab, "EGG," Available: https://unedvoicelab.com/egg/. Accessed June 23, 2024.

[61] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," Software available from tensorflow.org, 2015.

[62] F. Chollet *et al.*, "Keras," Available: https://github.com/keras-team/keras. Accessed May 30, 2024.

Mihai-Andrei Costandache

Received June 15, 2024
Revised June 24, 2024
Accepted June 27, 2024

Mihai-Andrei Costandache
"Alexandru Ioan Cuza" University of Iasi, Romania,
Faculty of Computer Science
General Berthelot, No. 16, Iasi, Romania
E–mail: `andrei97mihai@gmail.com`