VISHNIAKOU U.A. SHAYA B.H.

METHODICS AND TOOLS OF COUGH SOUND PROCESSING ON BASIC OF NEURAL NET

Belarusian state University of Informatics and Radioelectronics
Minsk, Republic of Belarus

The purpose of the article is to analyze the methods and means of processing cough sounds to detect lung diseases, as well as to describe the developed system for classifying and detecting cough sounds based on a deep neural network. Four types of machine learning and the use of convolutional neural network (CNN) are considered. Hypermarkets of CNN are given. Varieties of machine learning based on the CNN are discussed. The analysis of works on the methodology and means of processing cough sounds based on the CNN with the reduction of the means used and the accuracy of recognition is carried out. Details of machine learning using the environmental sound classification 50 (ESC-50) dataset are discussed. To recognize COVID-19 cough, a classifier was analyzed using CNN as a machine learning model. The proposed CNN system is designed to classify and detect cough sounds based on ESC-50. After selecting a set of sound classification data, four stages are described: extraction of features from audio files, labeling, training, testing. The ESC-50 used for the study was downloaded from the Kaggle website. Python libraries and modules related to deep learning and data science were used to implement the project: NumPy, Librosa, Matplotlib, Hickle, Sci-Kit Learn, Keras. The implemented network used a stochastic gradient algorithm. Several volunteers recorded their voices while coughing using their smartphones and it was assured to record their voices in a public environment to introduce noise to the sounds, in addition to some audio files that were downloaded online. The results showed an average accuracy of 85.37 %, precision of 78.8 % and a recall record of 91.9 %.

Keywords: *machine learning, neural network, cough sound processing, cough classification system.*

Introduction

Machine learning (ML) uses a data set, which is a table containing rows and columns, information about the classified object. The columns represent a characteristic of the object of interest, and each row represents one object. Predictions and pattern recognition using machine learning can be performed using several types [1].

1. Supervised: input of functions into the algorithm in addition to the result, which is called a class or label. Supervised machine learning is the most commonly used type. Along with this form, there are two main processes: the learning process, in which the algorithm receives big data and allows it to study the output data or result, and the testing phase, in which the remaining data is transmitted to the algorithm to obtain predictions and compare the result.

2. Unsupervised: unlike supervised learning, unsupervised learning is seen as an advanced way of machine learning. An unlabeled data set is passed to an algorithm that will perform its calculations and make predictions without any control.

3. Semi-supervised: a combination of both supervised and unsupervised learning, when algorithms can be assigned a labeled dataset, and training will take place. After that, according to their training, the algorithm continues for the remaining unlabeled datasets.

4. Reinforcement: learning from mistakes, machines can receive a set of instructions and templates that must act according to what is given in their path until the algorithm learns the optimal solution.

With neural networks, both supervised and unsupervised machine learning can be used. Training of a neural network with a labeled dataset can be performed by correcting the error and updating the weights using the back propagation method by moving backwards from the output layer through the hidden layers until the difference between the errors becomes zero. Unsupervised learning does not have a labeled dataset and depends on the categorization and ordering of the data.

Machine learning algorithms depend on some optimal parameters or hyper-parameters to train the model [2]. Hyper-parameters of NN are defined:

1. Learning rate: a variable that determines the speed of the learning process and parameter updates. The lower the learning rate, the slower the learning process.

2. Momentum: the direction of what will happen next is known by the momentum, which also helps to prevent fluctuations.

3. Batch size: each training iteration uses a certain number of samples, called the batch size.

4. Number of epochs: the number of iterations in which all input data is transmitted back and forth over the network during the learning process. The weights change during each epoch. The number of epochs should be increased until the model is well fitted.

5. The number of hidden layers: between the input and output layers determines the complexity of the network. The optimal number of hidden layers or blocks can be detected when the error remains constant.

In this article we consider the methodology and means of cough recognition using machine learning, neural networks, Python language tools.

Convolutional neural network

Convolutional Neural Network (CNN) has shown high performance in classification and detection in many areas, especially related to pattern recognition, such as image classification, video tagging and sound classification, etc. The term convolution was first mentioned by Lekun et al. [3], since then CNN has become more popular and attracted serious attention of researchers and programmers [3, 4]. CNN consists of input and output layers and three main hidden layers: convolutional, pulling, fully connected.

1. Convolutional layer: convolution is a mathematical operation defined as "an integral that expresses the amount of overlap of one function g when it is shifted by another function f" [3], used in a convolutional layer for unification.

2. Pulling layer: dimension reduction is performed inside this layer, this is also called downsampling. There are two types of this layer: maximum and average union.

3. Fully connected layer: before reaching the output layer, each neuron in the output layer connects to each node in this layer.

Sound processing

Encoding of audio signals can be represented in two main types: analog, and digital. Second is more widely used, in which the sound wave is sampled at a certain speed, includes MP3 and WAV types [5]. Audio objects are extracted according to audio data standards. They include the abstraction level (high, mid, low), the time domain (instantaneous, segment level, global), the signal domain (time domain, frequency, time-frequency). The article [6] discusses two types of sound functions in ML:

1. Traditional ML: The input data for the ML model are functions, including both time and frequency domains. An example of the most widely used functions is the Zero crossing rate (ZCR).

2. Deep learning: The approach is based on CNN and RNN methods for extracting signal features, using feature extraction methods known as spectrograms, Mel- spectrograms, Mel-frequency cepstrum coefficients (MFCC), perception linear prediction (PLP) and linear prediction coding (LPC).

a. PLP: Feroze et al. proposed a CNN classifier for detecting sound events in a polyphonic environment, where the PLP feature extraction method was used for transmission to the network [7]. The authors in [8] presented a comparative cough detection system. They encoded audio signals to visualize them as images and transmit them to CNN using 5 different PLP techniques. The RASTA-PLP spectral method showed best performance among others with an average accuracy of 0.99.

b. LPC: is considered a function of the time domain. Chowdhury et al. [9] proposed an approach to speaker recognition using MFCC and LPC functions, which are considered to be frequently used functions. These functions are combined with the use of 1D triplet CNN to improve speaker recognition performance.

c. MFCC: since audio signals are considered non-stationary, so that the values change rapidly over time, most feature extraction methods depend on short-term processing. MFCC are based on the perception of human hearing, since the perception of hearing takes into account the magnitude of the frequency components. Compared to perception of frequencies and magnitudes by human hearing, MFCC uses a nonlinear representation of magnitudes, frequency scaling using filter banks.

A deep CNN with two convolutional layers and two connected layers was trained and tested on the basis of MFCC with a low frequency for three different sets of environmental data in the work [10]. The results of the evaluation after cross-validation training using 10-folds and 5- folds validation showed accuracy for each data set of 64.5 %, 73.1 % and 73.7 %.

Cough detection methods

Bales et al. [11] presented a respiratory infection monitoring tool using CNN to detect cough sounds in order to diagnose bronchitis, bronchiolitis and pertussis. The environmental sound classification (ESC) dataset was used to feed to CNN, which was pre-processed, the lines were labeled as coughing and non-coughing with 993 samples for each. The raw audio signals from the dataset were converted into a Mel spectrogram, resulting in 432x288x3 images, which, in turn, were converted to grayscale. The CNN structure proposed to reduce its complexity consisted of 5 layers: a 2x2 max pooling layer, two convolutional layers, a ReLU activation function, and a 2x2 max pooling layer, in addition to the mel spectrogram image as an input and output layer. The results obtained after dividing the data set by 70-15-15 % in the quality of training, testing and verification showed an F1 score of 89.35 %.

The authors in [12] evaluated CNN for detecting cough sounds, where the sound sounds were segmented, and STFT was performed to obtain 64x16 spectrograms for a convolutional neural network, and labeled as coughing and non-coughing. CNN consists of five layers, 2 convolutional, 2 fully connected and a Softmax classification layer. The final data set contained 627 samples of cough sounds, also sampled up to 44.1 kHz and 16 kHz. Five experiments were conducted to evaluate the models. Comparative tables showed that the model was able to detect cough with a high accuracy of 82.5 %.

A diagnostic approach to pneumonia was proposed in [13] using logistic regression as a machine learning classifier. Mel Cepstral coefficients and non-Gaussianity index characteristics were added together with the extracted wavelet characteristics from the collected sounds. With the help of a microphone, a data set was collected from 91 patients suffering even from

pneumonia, asthma or bronchitis. Sensitivity of 94 % and specificity of 63 % were achieved using wavelet functions. However, after combining other features, the specificity increased to 88 %.

Using CNN, the authors in [14] proposed a cough detection model to prove the possibility of obtaining high performance using sounds recorded by mobile phones. 43 volunteers were asked to cough 20 times at two different distances using 5 different smart phones. The sampling rate was 44.1 kHz, and the data transfer rate was 16 bits. The final data set consisted of 6737 samples labeled as cough and 8854 labeled as control sounds. The results of the two implemented scenarios were evaluated and compared with the previous work, which showed an improvement in accuracy that reached 91.7 % using the CNN and Mel-spectrogram in the first scenario. They also recorded an accuracy of 86.7 % in the second scenario using the same algorithm.

Using Tensorflow, a deep learning library from Google, Khomsei et al. [15] presented a cough detection method using a deep learning network. Principal component analysis was implemented as a feature extraction method. The data set was collected with the help of eight volunteers to record cough sounds using a microphone connected to a raspberry pi, which has a sampling rate of 44.1 kHz. The final data set contained a total of 810 sounds, 229 of them are productive sounds, 74 are unproductive and finally 507 were sounds unrelated to coughing. The division of the training test was divided by 70/30 % in such a way that two types of experiments were used to create the model: the first is to evaluate the dataset using DLN only, and the second is using PCA with DLN.

Cough detection using the ESC 50 dataset

A machine learning approach using ESC-50 dataset was proposed in [16]. MFCC was first used as an object extraction method to extract 5000x36 objects, then PCA was applied to reduce the dimension to 1D vector 107. Random frog, UVE algorithm and VIP algorithm were used to reduce complexity and select informative features. After applying the feature selection methods, the SVM linear classifier was applied to classify sounds into coughing and non-coughing sounds. UVE with 20 functions showed the best performance results with an F1 score of 95 %.

ESC 50 in addition to two other datasets FSDKaggle2018 and Coughvid were evaluated in [17] in order to eliminate symptoms associated with breathing. CoughNet, the proposed model, takes input data in the form of a Mel-spectrogram and imports it into CNN. The model is applied to a CNN-LSTM processor that accepts records as input. The prototype is developed using Verilog HDL on the Xilinx Kintex-7 160t FPGA, which has a lower power consumption.

To classify the COVID-19 cough, Bansal et al. [18] implemented a cough classifier using CNN as a machine learning model. Audioset and ESC-50 were the two selected datasets to train their model, which are labeled COVID and non-COVID in addition to MFCC and spectrogram as feature extraction methods for the proposed model. The dataset contained 871 YouTube videos and 40 audio files, which were reduced to 501 audio files separated for training and testing. The CNN architecture consists of three convolutional layers and three fully connected layers. The overall accuracy of the resulting model was 70 %.

The proposed cough detection system based on CNN

The proposed system is designed to classify and detect cough sounds. After selecting a sound classification dataset, there are four main steps. The first step is to extract features from audio files, such as MFCCs, chromagram, Mel-spectrogram, spectral contrast and tonal centroid features. The second stage is the labeling stage, so we classify the sound samples into coughing and non-coughing, then we enter the input data into CNN. Then we proceed to the training stage and record the results until we reach the optimal parameters in accordance with the best indicators (change in the number of epochs, learning rate, etc.). At the final stage, after creating the model, several tests were applied to the recorded sounds from the volunteers (Figure 1).
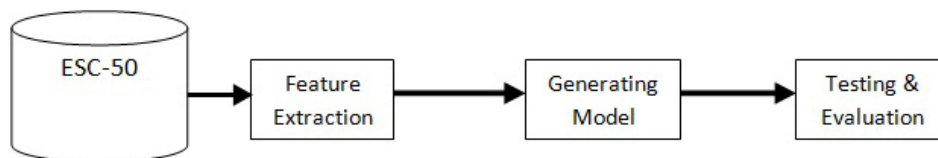

Figure 1. Stages of the proposed system

The ESC dataset [19] is a set of 2,000 sound files representing 50 types of environmental sounds. ESC-50 was published in 2015 and has been used in many publications and systems. ESC-50 was used for research and downloaded from Kaggle's website for data scientists, which contains over 50,000 datasets in addition to over 400,000 publicly available notebooks.

Software environment for implementation

More than 60 % of developers use Python because of its simplicity and a wide range of useful libraries, such as Pandas, Sci-Kit learn, Matplotlib, Seaborn and others. Tensorflow [20], an open source ML platform developed by Google for deep machine learning. Google Colab, designed for Python programmers. Colab specializes in data science and ML, users can import datasets from Kaggle. For our project, the following Python libraries and modules were used, which are related to deep machine learning and data science.

1. NumPy is a library used for structuring arrays, manipulating shapes, sorting. Several functions from NumPy were used, such as the "absolute" function, which was used with the Librosa library to extract short-term fourier transform (STFT) features from audio files.

2. Librosa is an indispensable library used for audio, speech and music analysis. Using Librosa, six methods of feature extraction were applied, starting with STFT, then MFCC and Mel-spectrogram, the remaining three methods were spectral contrast, chroma and tonnetz.

3. Matplotlib: for plotting and visualizing arrays, data and statistics. We used it to construct the matrix of the system and related graphs.

4. Hickle is used to save functions extracted from audio files in HDF5 type format.

5. Sci-Kit Learn: used to implement machine learning algorithms, including supervised and unsupervised. SK-learn was used to obtain model results and metrics like accuracy and recall.

6. Keras: to implement a deep learning model. The training, testing and evaluation of our system is based on Keras and its built-in functions.

Feature extraction

Since our dataset is audio data, we used the Librosa library to extract functions, characteristics (STFT, MFCC, Mel-spectrogram, chromati, spectral contrast and tonnetz) and analyze them. 193 features as total were extracted using Librosa from all the above mentioned methods and resulted in a data-frame that contained 1977 rows and 193 columns from all the input data The train function accepts the following parameters: objects, labels, type, number of classes, epochs, and optimizer.

The neural network adjusts its weights according to the results obtained after each iteration using various algorithms known as optimizers, which implies calculating the difference between the results. The implemented network uses the stochastic gradient decent as an optimizer for its simplicity, which calculates the gradient of the loss function in the network. The initial learning rate for this system was 0.1 and momentum = 0.9. The functions are those that were saved using Hickle, as well as labels. The number of epochs usually needs to be changed to get the best results, we have increased the number of epochs from 300 to 500, 750, 1200, 1500.

To predict, some sounds were tested and classified using the "predict" function in addition to some recorded cough sounds of volunteers. The prediction method was implemented using a loop to iterate through several audio files and output the accuracy of the detected result for the top three classes that received the highest results. Using the Librosa and Matplotlib libraries, the extracted objects were presented as images showing the different forms of each extraction method.

Several volunteers recorded their voices while coughing using their smart phones and it was assured to record their voices in a public environment to introduce noise to the sounds, in addition to some audio files that were downloaded online. The results showed an average accuracy of 85.37 %, precision of 78.8 % and a recall record of 91.9 %.

Conclusion

1. This thesis introduces the state of art of the most relevant cough detection systems and discuss their obtained results. As well, a well-designed cough detection system was build using the public dataset ESC-50, that was fed a convolutional neural network. The layers of the network as well as the training methods were stated in details.

2. After generating the model, another sound set was tested in order to evaluate the model. Thus, several volunteers recorded their voices while coughing using their smart phones and it was assured to record their voices in a public environment to introduce noise to the sounds, in addition to some audio files that were downloaded online. The results showed an average accuracy of 85.37 %, precision of 78.8 % and a recall record of 91.9 %.

REFERENCES

1. **Radhakrishnan P. Towards Data Science. 9 Aug 2017.** [Electronic resource]. Avialable : https://towardsdatascience.com/what-are-hyperparameters-and-how-to-tune-the-hyperparameters-in-a-deep-neural-network-d0604917584a. Date of access: Aug 2022.

2. **LeCun J., Boser B., Denker J., Henderson D., Howard R., Hubbard W., Jackel L.** Backpropagation Applied to Handwritten Zip Code. Neural Computation, 1989. – Vol. 1, No. 4. – Pp. 541-551.

3. **Dhillon A., Verma G.K.** Convolutional neural network: A review of models, methodologies and applications to object detection. Progress in Artificial Intelligence, 2019. – Vol. 9, No. 2. – Pp. 85-112.

4. **Kriszhevsky G.K., Sutskever G.K., Hinton G.E.** ImageNet classification with deep convolutional neural networks // Proceedings of the 25th International Conference on Neural, vol. 60, No. 6, 2012. – Pp. 1097-1105.

5. **Chu, S.** Unstructured Audio Classification for Environment Recognition.// Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, 2008. – Pp. 1845-1846.

6. **Banchhor K.S., Khan A.** Musical Instrument Recognition using Zero Crossing Rate and Short-time Energy. International Journal of Applied Information Systems. vol. 1, No 3, 2012. – Pp. 85-97.

7. **Feroze K., Maud A.R.** Sound Event Detection in Real Life Audio using Perceptual Linear Predictive Feature with Neural Network // Proceeding of 15th International Bhurban Conferance on Applied Sceinces & Technology, Islamabad, Pakistan, 2018.

8. **Wang H.-H., Liu J.-M, You M., Li G.-Z.** Audio Signals Encoding for Cough Classification Using Convolutional Neural Networks: A Comparative Study // IEEE International Conference on Bioinformatics and Biomedicine, Shanghai, 2015.

9. **Chowdhury A., Ross A.** Fusing MFCC and LPC Features using 1D Triplet CNN for Speaker Recognition in Severely Degraded Audio Signals. IEEE Transactions on Information Forensics and Security, vol. 15, 2019. – Pp. 1616-1629.

10. **Piczak K.J.** Environmental sound classification with convolutional neural networks // IEEE International workshop on machine learning for signal processing, Boston, 2015.

11. **Bales C., Nabeel M., John N.C., Masood U., Qureshi H., Farooq H.** Can Machine Learning Be Used to Recognize and Diagnose Coughs? // The 8th IEEE International Conference on E-Health and Bioengineering, Romania, 2020.

12. **Amoh A., Odame K.** Deep Neural Networks for Identifying Cough Sounds. IEEE Transactions on Biomedical Circuits and Systems, vol. 10, No. 5, 2016. – Pp. 1003-1011.

13. **Kosasih K., Abeyratne U., Swarnkar V., Triasih R.** Wavelet Augmented Cough Analysis for Rapid Childhood Pneumonia Diagnosis. IEEE Transactions on biomedical engineering, vol. 62, No. 4, 2015. – Pp. 1185-1194.

14. **Barata F., Kipfer K., Weber M., Tinschert P., Fleisch E., Kowatsch T.** Towards Device-Agnostic Mobile Cough Detection with Convolutional Neural Networks // IEEE International Conference on Healthcare Informatics, Xi'an, 2019.

15. **Khomsay S., Vanijjirattikhan R., Suwatthikul J.** Cough detection using PCA and Deep Learning // Intern. Conf. on Information and Communication Technology Convergence (ICTC), Jeju, 2019.

16. **Chen X., Hu X. Zhai G.** Cough Detection Using Selected Informative Features from Audio Signals. Cirnell University, August, 2021.

17. **Rashid H.-A., Mazumder A.N., Pan U., Niyogi K., Mohsenin T.** CoughNet: A Flexible Low Power CNN-LSTM Processor for Cough Sound Detection // IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS), Washington DC, 2021.

18. **Bansal V., Pahwa G., Kannan N.** Cough Classification for COVID-19 based on audio mfcc features using Convolutional Neural Networks // IEEE International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, 2020.

19. **Piczak K.J.** ESC: Dataset for Environmental Sound Classification // Proceedings of the 23rd Annual ACM Conference on Multimedia, Brisbane, 2015.

20. **Abadi M., Barham P., Chen P., Chen Z, Davis A.** TensorFlow: A System for Large-Scale Machine Learning. Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16). November, 2016 Savannah, GA, USA.

ЛИТЕРАТУРА

1. **Radhakrishnan P. Towards Data Science. 9 Aug 2017.** [Electronic resource]. Avialable : https://towardsdatascience. com/what-are-hyperparameters-and-how-to-tune-the-hyperparameters-in-a-deep-neural-network-d0604917584a. – Data of accsses: Aug 2022.

2. **LeCun J., Boser B., Denker J., Henderson D., Howard R., Hubbard W., Jackel L.** Backpropagation Applied to Handwritten Zip Code. Neural Computation, vol. 1, No. 4, 1989. – Pp. 541-551.

3. **Dhillon A., Verma G.K.** Convolutional neural network: A review of models, methodologies and applications to object detection. Progress in Artificial Intelligence, vol. 9, No. 2, 2019. – Pp. 85-112.

4. **Kriszhevsky G.K., Sutskever G.K., Hinton G.E.** ImageNet classification with deep convolutional neural networks // Proceedings of the 25th International Conference on Neural, vol. 60, No. 6, 2012. – Pp. 1097-1105.

5. **Chu, S.** Unstructured Audio Classification for Environment Recognition // Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, 2008. – Pp. 1845-1846.

6. **Banchhor K.S., Khan A.** Musical Instrument Recognition using Zero Crossing Rate and Short-time Energy. International Journal of Applied Information Systems. vol. 1, N 3, 2012. – Pp. 85-97.

7. **Feroze K., Maud A.R.** Sound Event Detection in Real Life Audio using Perceptual Linear Predictive Feature with Neural Network // Proceeding of 15th International Bhurban Conferance on Applied Sceinces & Technology, Islamabad, Pakistan, 2018.

8. **Wang H.-H., Liu J.-M., You M., Li G.-Z.** Audio Signals Encoding for Cough Classification Using Convolutional Neural Networks: A Comparative Study // IEEE International Conference on Bioinformatics and Biomedicine, Shanghai, 2015.

9. **Chowdhury A., Ross A.** Fusing MFCC and LPC Features using 1D Triplet CNN for Speaker Recognition in Severely Degraded Audio Signals. IEEE Transactions on Information Forensics and Security, vol. 15, 2019. – Pp. 1616-629.

10. **Piczak K.J.** Environmental sound classification with convolutional neural networks // IEEE International workshop on machine learning for signal processing, Boston, 2015.

11. **Bales C., Nabeel M., John N.C., Masood U., Qureshi H., Farooq H.** Can Machine Learning Be Used to Recognize and Diagnose Coughs? // The 8th IEEE International Conference on E-Health and Bioengineering, Romania, 2020.

12. **Amoh A., Odame K.** Deep Neural Networks for Identifying Cough Sounds. IEEE Transactions on Biomedical Circuits and Systems, vol. 10, No. 5, 2016. – Pp. 1003-1011.

13. K**osasih K., Abeyratne U., Swarnkar V., Triasih R.** Wavelet Augmented Cough Analysis for Rapid Childhood Pneumonia Diagnosis. IEEE Transactions on biomedical engineering, vol. 62, No. 4, 2015. – Pp. 1185-1194.

14. **Barata F., Kipfer K., Weber M., Tinschert P., Fleisch E., Kowatsch T.** Towards Device-Agnostic Mobile Cough Detection with Convolutional Neural Networks // IEEE International Conference on Healthcare Informatics, Xi'an, 2019.

15. **Khomsay S., Vanijjirattikhan R., Suwatthikul J.** Cough detection using PCA and Deep Learning // Intern. Conf. on Information and Communication Technology Convergence (ICTC), Jeju, 2019.

16. **Chen X., Hu X. Zhai G.** Cough Detection Using Selected Informative Features from Audio Signals. Cirnell University, August, 2021.

17. **Rashid H.-A., Mazumder A.N., Pan U., Niyogi K., Mohsenin T.** CoughNet: A Flexible Low Power CNN-LSTM Processor for Cough Sound Detection // IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS), Washington DC, 2021.

18. **Bansal V., Pahwa G., Kannan N.** Cough Classification for COVID-19 based on audio mfcc features using Convolutional Neural Networks // IEEE International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, 2020.

19. **Piczak K.J.** ESC: Dataset for Environmental Sound Classification // Proceedings of the 23rd Annual ACM Conference on Multimedia, Brisbane, 2015.

20. **Abadi M., Barham P., Chen P., Chen Z, Davis A.** TensorFlow: A System for Large-Scale Machine Learning. Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16). November, 2016 Savannah, GA, USA.

ВИШНЯКОВ В.А., САЙЯ Б.Х.

МЕТОДЫ И СРЕДСТВА ОБРАБОТКИ ЗВУКОВ КАШЛЯ НА БАЗЕ НЕЙРОННОЙ СЕТИ

*УО «Белорусский государственный университет информатики и радиоэлектроники»
г. Минск, Республика Беларусь*

Целью статьи является анализ методики и средств обработки звуков кашля для выявления легочных заболевания, а также описание разработанной системы для классификации и обнаружения звуков кашля на базе глубинной нейронной сети. Рассмотрены четыре типа машинного обучения, использование сверточной нейронной сети (СНС). Приведены гипермаркеты СНС. Обсуждены разновидности машинного обучения на базе СНС. Выполнен анализ работ по методике и средствам обработки звуков кашля на базе СНС с приведением используемых средств и точности распознавания. Обсуждены детали машинного обучения с использованием набора данных классификации звуков (environtmental sound classification 50 – ESC-50). Для распознования кашля COVID-19 проанализирован классификатор, используя CNN в качестве модели машинного обучения. Предлагаемая система СНН предназначена для классификации и обнаружения звуков кашля на базе ESC-50. После выбора набора данных классификации звуков описаны четыре этапа: извлечение признаков из аудиофайлов, маркировки, обучение, тестирование. ESC-50, использованный для исследования, был загружен с веб-сайта Kaggle. Для реализации проекта были использованы библиотеки и модули Python, которые связаны с глубоким обучением и наукой о данных: NumPy, Librosa, Matplotlib, Hickle, Sci-Kit Learn, Keras. Реализованная сеть использовала алгоритм стохастического градиента. Несколько добровольцев записали свои голоса во время кашля с помощью своих смартфонов, было гарантировано, что они будут записывать свои голоса в общественных местах, чтобы внести шум в звуки, в дополнение к некоторым аудиофайлам, которые были загружены онлайн. Результаты показали среднюю точность 85,37 %, точность 78,8 % и рекорд отзыва 91,9 %.

Ключевые слова: машинное обучение, нейронные сети, обработка звуков кашля, система классификации кашля.

Vishniakou Uladzimir, doctor of technical science, professor of ICT department of Belarusian State University of Informatics and Radioelectronics. Research interest: information management and security, electronic business, intellectual management systems. Mingled of two doctoral counsels of thesis's defense. Author more 400 scientific publications including 6 monographs (1 - English), 4 study books with stamp of education Ministry, 8 volumes manual «Information management», 175 scientific articles.

Вишняков Владимир Анатольевич – д.т.н., профессор профессор БГУИР, каф. ИКТ. Область научных интересов: информационное управление и безопасность, электронный бизнес, интеллектуальные системы управления. Член 2-х докторских Советов по защите диссертаций. Автор более 400 научных работ, в том числе 6 монографий (1 на английском языке), 4-х учебных пособий с грифом Министерства образования, 8-и томного учебного комплекса «Информационный менеджмент», 175 научных статей.

E-mail: vish2002@list.ru

Shaya Bahaa, master of technical science, PhD-student of ICT department of Belarusian State University of Informatics and Radioelectronics.

Сайя Бахаа – магистр технических наук, аспирант кафедры ИКТ БГУИР.