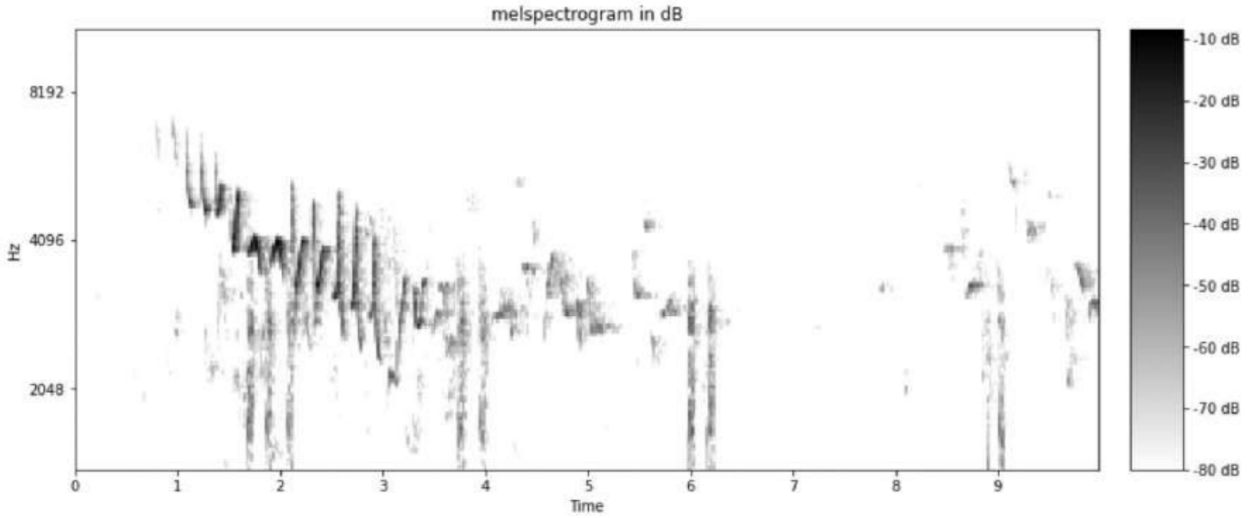

Рис. 1: A spectrogram of preprocessing the voice of a Sylvia Atricapilla bird (Lieska-cornahalouka)

S_ms_db = librosa.amplitude_to_db (np.abs(S_ms), ref=np.max)

Where parameters are

sr=22050, n_fft=1024, hop_length=512, n_mels=128.

An example of preprocessing the voice of a Sylvia Atricapilla bird (Lieska-cornahalouka) is a high-pass filter, calculation of mel frequency cepstral coefficients and display as a spectrogram on a graph is shown on figure 2.

VI. Model architecture

Having a Mel spectrogram of the bird's sounds, we will recognize its species among a pre-prepared list of species (classes). The model extracts specific characteristics of the processed data (images), sends them to the input of a deep neural network, and outputs a set of probabilities that correspond to the likelihood that the image belongs to each of these classes. We assume that the class with the highest of these probabilities is the output of the model. This is a deep neural network of the CNN type (Convolutional Neural Network - a convolutional neural network) for recognizing the image class of birds' voices. It is a type of deep learning algorithm that is commonly used in image and video processing applications. Convolutional neural networks are specifically designed to process data that has a grid-like topology, such as images, where the goal is to classify an image into one of several categories. They work by applying a series of filters or convolutions to the input data, which extract features from the input data. These features are then used to make predictions about the input data. They are also used in object detection to locate objects within an image, and in image segmentation for partitioning an image into regions or segments. Overall, CNNs have revolutionised the field of computer vision and have led to significant improvements in image and video processing applications.

In the architecture, we used the EfficientNetB3 network as well as three more layers:

(*Flatten, Dropout, Dense with the softmax function as an output*) to build a CNN network. The following network structure is EfficientNetB3 Flatten Dropout Dense(with softmax function as output).

EfficientNet (https://arxiv.org/abs/1905.11946, https://keras.io/api/applications/efficientnet/efficientnetb3-function) is a modern development of a convolutional neural network from Google Brain (see figure 3). The main objective of EfficientNet was to thoroughly test how to scale the size of convolutional neural networks. For example, one can scale ConvNet based on layer width, layer depth, input image size, or a combination of all these parameters. Thus, the final model was built on the basis of EfficientNetB3 and 14 different classes (bird species) with Adam optimizer, categorical cross-entropy loss function and balanced class weights.

VII. Quality of model training

The quality of bird voice recognition by the represented model is 75.6 percent of the total accuracy on the test data, where

- 7 classes have an F1-score of more than 80 percent
- 3 classes have an F1-score between 70 percent and 80 percent
- 2 classes have an F1-score between 60 percent and 70 percent
- 2 classes have F1-score less than 60 percent.

Moreover, a prototype of the model for automated bird voice recognition "*Bird Sound Recognizer*"was created for the implementation of autonomous continuous monitoring of rare threatened species, indicator species and the state of biodiversity in forest ecosystems [4]. It is located on the online platform corpus.by (http: https://corpus.by/BirdSoundsRecognizer/?lang=en) and is open and free for using [5].

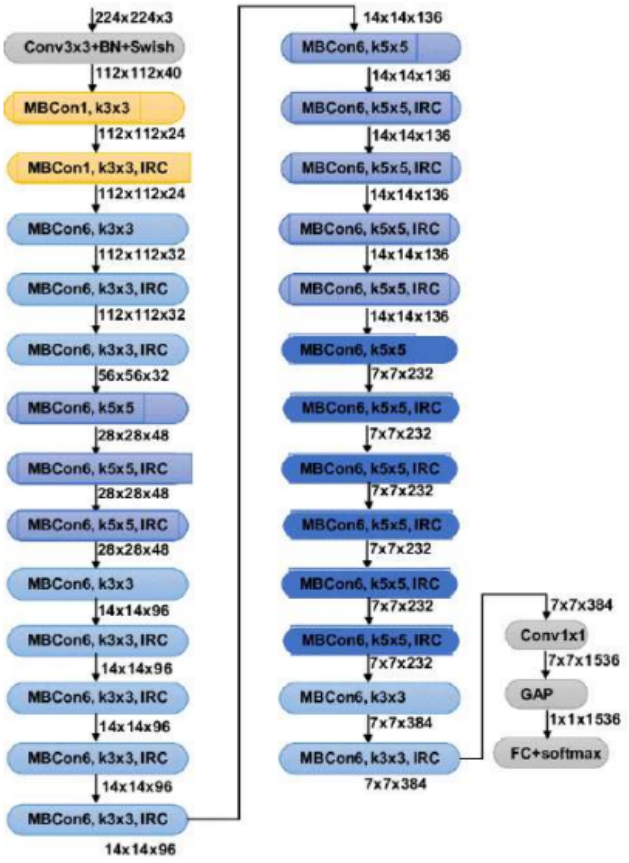

Figure 3. EfficientNet Network architecture

VIII. Model Improvements Based on Ontological Approaches

One of the ways to further improve the quality of the model is the ability to take into account, in addition to the sound signal itself, various meta-information: place, time, habitat of a particular species of birds. For a comprehensive solution of this problem, allowing to take into account the various features of the subject area, it is proposed to use the ontological approach.

When it comes to recognizing birdsong from an audio signal, ontological approaches are essential for several reasons. First and foremost, ontologies provide a way to organize and structure the vast amounts of data that are available for each uploaded audio file. In this case, the meta-information available for each audio file can be used to create a taxonomy of bird species, subspecies, and groups. This makes it easier to organize and analyze the data, and can help to improve the accuracy of the recognition process.

In addition, ontological approaches can help to address issues related to data inconsistency and incompleteness. By using a well-defined and structured ontology, it becomes easier to identify and correct inconsistencies in the data, such as misspellings or variations in naming conventions.

Furthermore, ontologies can help to facilitate the integration of different data sources, such as data from different recording locations or from different researchers. By using a common ontology, it becomes easier to merge and compare data from different sources, which can help to improve the accuracy of the recognition process.

Finally, ontological approaches can help to support the development of intelligent systems that are capable of learning and adapting to new data. By using a well-defined ontology as a basis for machine learning algorithms, it becomes possible to create systems that can recognize new bird species and adapt to new recording conditions.

The choice of technology for implementing the ontological approach is also important. Because it should provide a sufficient level of flexibility and scalability to integrate various types of systems and knowledge into them. As a technological basis, it is proposed to use the OSTIS technology, the main advantages and specifics of which in signal processing and integration with DNN are presented in papers [6-10]. An example of a fragment of a top-level ontology using OSTIS sc-code is presented below.

Sylvia Atricapilla
:= [Eurasian blackcap]
∈ *English language*
:= [*Lieska – cornahalouka*]
∈ *Belarusian language*
∈ *specie*
⊂ *sylvia*
∈ *genus*
⊂ *old world warbler*
∈ *family*
⊂ *perching birds*
∈ *order*
⊂ *vertebrates*
∈ *phylum*
⊂ *animal*
∈ *kingdom*
→ *habitat**:
{• *Eastern Europe*
→ *coordinates**:
{• [*From*: 54.1343° *N* 28.5079°*E*]
• [*To*: 54.5028°*N*, 28.8794° *E*]
• *Western Asia*
• *Northwestern Africa*
}

By using a structured and well-defined ontology to organize and analyze the available data, it becomes possible to create intelligent systems that are capable of accurately recognizing a wide range of bird species, subspecies, and groups.

IX. Conclusion

The article describes a model for recognizing

the voices of Belarusian birds. It is based on the analysis of a melfrequency cepstrum (MFCC, mel-frequency cepstrum). The Mel spectrogram is a graphical representation of an audio signal in which frequencies are represented on Mel scale instead of the linear frequency scale used in a conventional spectrogram. For machine learning of the model, a deep neural network of the CNN (Convolutional Neural Network) was used to recognize the image class of birds' voices, since this type of network is more suitable for image recognition tasks. To build the CNN network, we chose the EfficientNetB3 network, as well as three more layers (Flatten, Dropout, Dense with the softmax function as output). Thus, the final model was built on the basis of EfficientNetB3 and 14 different classes (bird species) with Adam optimizer, categorical cross-entropy loss function and balanced class weights.

The overall recognition quality of the model was 75.6 percent. The conduction of the experiment was fulfilled on 14 species of birds with 200 records for each of the species using a CNN-based model with spectrograms as a characteristic of the input signal to the model. In the future, it is planned to expand the list of bird species for recognition to 116. The next step of the project is to monitor the continuous signal for the detection of the bird's voice in real time.

To solve this task, a recognition system will be designed using another data set, the audio files of which are annotated in detail by ornithologists, taking into account the time stamps for the bird's voice.

In paper also proposed approaches to further improve the quality of ML-models through the use of the ontological approach and OSTIS technology.

REFERENCES

[1] F. Briggs, R. Raich, X. Z. Fern Audio Classification of Bird Species: A Statistical Manifold Approach. ICDM 2009, The Ninth IEEE International Conference on Data Mining, Miami, Florida, USA, 6-9 December 2009, pp. 51–60.

[2] D. Stowell, M. D. Plumbley An open dataset for research on audio field recording archives: freefield1010. ArXiv, 2013, vol. abs/1309.5275.

[3] D. Stowell, M. D. Plumbley Automatic large–scale classification of bird sounds is strongly improved by unsupervised feature learning. *PeerJ*, 2014, vol. 2.

[4] S. A.Hajdurau, D. I. Latysevic, A. A. Bakunovic, L. I. Kajharodava, V. A. Chachlou, Ja. S. Zianouka, Ju. S. Hiecevic Madel baz danych dlia technalohii autamatyzavanaha raspaznavannia halasavych sihnalau zyviol [Database model for the technology of automated recognition of animal voice signals]. XXI International Scientific and Technical Conference "Development of Informatization and State System of Scientific and Technical Information RINTI-2022 UIIP NASB, Minsk, 2022, pp. 236–240.

[5] J. S. Hiecevic, J.S.Zianouka, A. S. Trafimau, A. A. Bakunovi c, D. I. Latyševic, A. JA. Drahun, M. M. Sliesarava, M. S. Tukaj Komplieks srodkau realizacyi zadac štucnaha inteliektu dlia bielaruskaj movy [A complex of means to implement tasks of artificial intelligence for the Belarusian language]. *Piervaja vystavka-forum IT-akadiemhrada Iskusstviennyj intielliekt v Bielarusi.* UIIP NASB. Minsk. 2022. P. 64-73. (In Belarusian).

[6] V. V. Golenkov, N. A. Gulyakina, D. V. Shunkevich Open technology of ontological design, production and operation of semantically compatible hybrid intelligent computer systems, Bestprint, Minsk, 2021, P. 690

[7] V. Golovko and et el. Integration of artificial neural networks and knowledge bases *Open Semantic Technologies for Intelligent Systems (OSTIS-2018)*, BSUIR, Minsk, 2018, pp. 133–146.

[8] M. Kovalev, A. Kroshchanka, V. Golovko, Convergence and integration of artificial neural networks with knowledge bases in next-generation intelligent computer systems *Open Semantic Technologies for Intelligent Systems (OSTIS-2022)*, BSUIR, Minsk, 2022, pp. 173–186.

[9] V. Zahariev, E. Azarov, K. Rusetski.An approach to speech ambiguities eliminating using semantically-acoustical analysis *Open Semantic Technologies for Intelligent Systems (OSTIS-2018)*, BSUIR, Minsk, 2018, pp. 211–222.

[10] V. Zahariev, K.Zhaksylyk, D. Likhachov, N. Petrovsky, M. Vashkevich, E. Azarov. Audio interface of next-generation intelligent computer systems *Open Semantic Technologies for Intelligent Systems (OSTIS-2022)*, BSUIR, Minsk, 2022, pp. 239–250.

Разработка системы распознавания звуков птиц с использованием онтологического подхода

Зеновко Е., Белявский Д., Кайгородова Л., Трофимов А., Хохлов В., Гецевич Ю., Захарьев В., Жаксылык К.

В работе предложена модель распознавания голосов птиц Республики Беларусь, основанная на анализе мел-спектрограмм (MFCC, mel-frequency cepstrum). Мелспектрограмма — это графическое представление звукового сигнала, в котором частоты представлены в мел-шкале вместо линейной шкалы частот, используемой в обычной спектрограмме. Шкала Mel — шкала высоты звуков, отсеивающая частоты звуков, которые человек не слышит, и оставляет самые характерные, находящихся на одинаковой дистанции для слушателя. Для машинного обучения модели была использована глубокая нейронная сеть типа CNN (Convolutional Neural Network) для распознавания класса изображения голоса птиц, так как именно этот вид сети больше подходит для задач распознавания изображений. Для построения сети CNN мы применили сеть EfficientNetB3, а также еще три слоя (Flatten, Dropout, Dense с функцией softmax в качестве выхода). Таким образом, окончательная модель была построена на основе EfficientNetB3 и 14 различных классов (видов птиц) с оптимизатором Адама (Adam optimizer), категориальной функцией потерь перекрестной энтропии (categorical cross-entropy loss function) и сбалансированными весами классов.

При проведении данного эксперимента на 14 видах птиц с 200 записями для каждого из видов и использованием модели на базе CNN со спектрограмами в качестве характеристики сигнала для входа на модель, получено общее качество распознавания 75,6 процентов. В дальнейшем планируется расширение списка видов птиц для распознавания до 116.

Показаны возможности использования онтологических подходов и технологии OSTIS для дальнейшего повышения качества моделей машинного обучения.

Received 13.03.2023