Developing an AI-Powered Bird Sound Recognition System for Monitoring Avian Biodiversity in Belarus

Yauheniya Zianouka, Volga Dydo, Alexandra Dydo, Darya Zhalova, Juras Hetsevich
United Institute of Informatics Problems
National Academy of Sciences of Belarus
Minsk, Belarus
{evgeniakacan, olgadydo1, daryazhalova, yuras.hetsevich}@gmail.com

Vadim Zahariev, Veronika Krischenovich
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
{zahariev, krish}@bsuir.by

***Abstract*—This paper explores the application of artificial intelligence (AI) of automated bird vocalisation recognition for the purpose of continuous monitoring of avian species in Belarus. It presents a novel system comprising a large, annotated dataset of Belarusian bird sounds, a convolutional neural network model trained for multi-label species identification, and a publicly available software platform. The AI-driven approach of the system enables efficient analysis of large audio datasets and real-time species identification. This research demonstrates the feasibility of using AI to overcome challenges in biodiversity monitoring and lays the groundwork for future advancements in automated wildlife conservation.**

***Keywords*—artificial intelligence, bird voice recognition, automatic processing, annotation, Mel spectrogram, convolutional neural network, deep learning, ontological approach.**

I. Introduction

Bird vocalisations are a critical component of avian behavior, ecology, and communication. The ability to recognize and interpret these vocal signals is essential for various fields, including ornithology, ecology, and conservation biology. Birds are among the most vocal animals, using a complex array of sounds for communication, mating, territory establishment, and alarm signaling. Understanding these vocalisations provides information on avian behavior and ecology, making it a vital area of study for researchers. As global biodiversity faces unprecedented threats, recognition of bird voices can play a crucial role in monitoring populations and informing conservation strategies.

The recognition of different bird species is complicated by numerous variations in their sound signals. Engaging ornithologists for manual processing of individual signals requires additional time, which significantly hinders the automation of species identification. This highlights the need to develop automated bird voice recognition systems capable of analysing large datasets [1]. Currently, there are several well-known applications, such as *Aubio, Prodigy, BirdNET, Merlin Bird ID by Cornell Lab, Song Sleuth, Warblr, and Chirpomatic*, that interpret bird vocalisations in various regions of the world [2]. The urgent need for a bird voice recognition system in Belarus arises from the fact that existing applications do not encompass the full biodiversity of birds in the country, with only a few capable of recognising European species. Belarus is home to a rich diversity of bird species, many of which are unique to the region or are part of migratory routes. Effective monitoring of these species is crucial for conservation efforts, especially for those that are rare or threatened. Traditional methods of monitoring birds often rely on manual observation and identification, which can be time-consuming and labor-intensive. Automated sound recognition systems can streamline this process by allowing the rapid analysis of large volumes of audio data. This efficiency enables researchers and conservationists to gather more comprehensive data on bird populations and their behaviors. Although there are several bird sound recognition applications available, many do not cover the full range of species found in Belarus. Developing a localized system ensures that unique vocalisations of Belarusian birds are accurately recognised and documented [3]. This specificity is crucial for effective conservation and research efforts tailored to the region's ecological needs.Furthermore, automatic voice recognition poses a particularly pressing challenge in countries that lack the resources to effectively monitor and protect wildlife. The successful development of such systems could have significant practical applications, including improved monitoring of bird populations, assessment of their ecological status, and conservation of biodiversity [4].

Therefore, the objective of the research is the development of a technology of automatic unmasking of residential signals to provide continuous automatic monitoring of rare, threatening, and other types of life and the state of vulnerability in forest ecosystems [5].

II. The Development of Bird Sound Recognition System Based on Artificial intelligence

Within the framework of the project, several tasks have been established:

1) Preparation of methodological foundations for the collection, annotation, and recognition of animal voice signals in Belarus.
2) Organizing data sets of existing electronic recordings of animal vocalisations from available sources.
3) Development of a structural framework for the automated recognition of animal voice signals to facilitate the the continuous monitoring of rare, threatened, and indicator species.
4) Creation of an analytical center for the continuous autonomous monitoring of rare, threatened, and indicator animal species.
5) Generating automated annotations of animal voice signals within the compiled databases.

In a previous paper [6] we described a dataset of bird vocalizations, the algorithms for their annotation, and the initial recognition model based on artificial intelligence (AI). The development of such systems is significantly enhanced by the application of AI for several reasons. Using machine learning algorithms, it can analyse complex sound patterns with a high degree of accuracy. Convolutional neural networks (CNN) learn to differentiate between subtle variations in bird vocalizations, leading to a more precise identification of species compared to traditional methods [7]. AI systems process large amounts of audio data quickly and efficiently. This scalability allows researchers to analyse extensive datasets collected from various locations over long periods, making it feasible to monitor large populations and diverse habitats without the need for extensive manual labour. AI models are trained and improved over time as more data become available. This ability to learn from new examples means that the system can adapt to changes in bird vocalisations, such as variations due to environmental factors or changes in species behavior, leading to ongoing improvements in recognition accuracy. AI systems can be designed to analyse audio streams in real-time, enabling immediate identification of bird species. This capability is particularly useful for field studies and monitoring programs, where timely data can inform conservation actions or research findings. AI can integrate audio data with other ecological data, such as weather conditions, habitat types, and geographic information. This holistic approach allows for more comprehensive analyses and a better understanding of the factors influencing bird populations and behaviors.

As part of preparing methodological foundations for the collection, annotation, and generation of animal vocal signals in Belarus, a comprehensive database was created from various sources. This database contains more than 2,500 audio recordings representing 116 bird species, which yield approximately 42,000 annotated vocalisations. The data set includes recordings of varying quality and audio features with different numbers of species that sing simultaneously. The annotation process for audio files intended for automated recognition involves segmenting them into appropriately sized windows, performing computer analysis, and generating a Mel spectrogram for each file (fig. 1). A Mel-spectrogram is a visual representation of sound that displays how the energy of a signal is distributed across different frequencies over time. It is derived from the Fourier transform, which analyses sound waves to extract their frequency components. In a Mel-spectrogram, the x-axis represents time, the y-axis represents frequency, and the intensity of colour or brightness indicates the amplitude (energy) at each frequency at each moment in time. The Mel scale is a perceptual scale of pitches that approximates the way humans perceive sound, making it particularly useful for audio analysis. The Mel scale emphasises frequencies that are more relevant to human hearing, allowing for better feature extraction that aligns with how we perceive sounds, which can improve the performance of recognition algorithms.

To create experimental software for the automatic recognition of animal vocal signals, efforts have been made to develop a model and its corresponding algorithms. This model is built on a pretrained convolutional neural network that incorporates additional layers, including *Flatten, Dropout, and Dense*. The activation functions for the final layer are *Softmax and Sigmoid*, with binary cross-entropy used as loss function. The task has evolved from Multi-Class Classification (one window – one prediction) to Multi-Label Classification (one window – multiple predictions). The neural network training process is carried out over 50 epochs, with adjustments to the learning rate based on the *ReduceLROnPlateu method*.

The input data is presented in the form of audio recordings and text annotations containing timestamps of the appearance of bird species. The refinement process includes three key steps (fig. 2). At the first stage, audio recordings are protected from background noises such as rain, wind, urban sounds, and other unwanted acoustic interference. For this purpose, the threshold spectrum processing method is used, which analyses the amplitude characteristics of the frequency components of the signal and drowns out the noise components, the level of which is below a given threshold. In the second stage, the audio signal is divided into 2-second windows according to the timestamps of the annotations. This process ensures the formation of a training sample in which each fragment contains bird sounds in a given time range. At the third stage, Mel-spectrograms are calculated for each selected

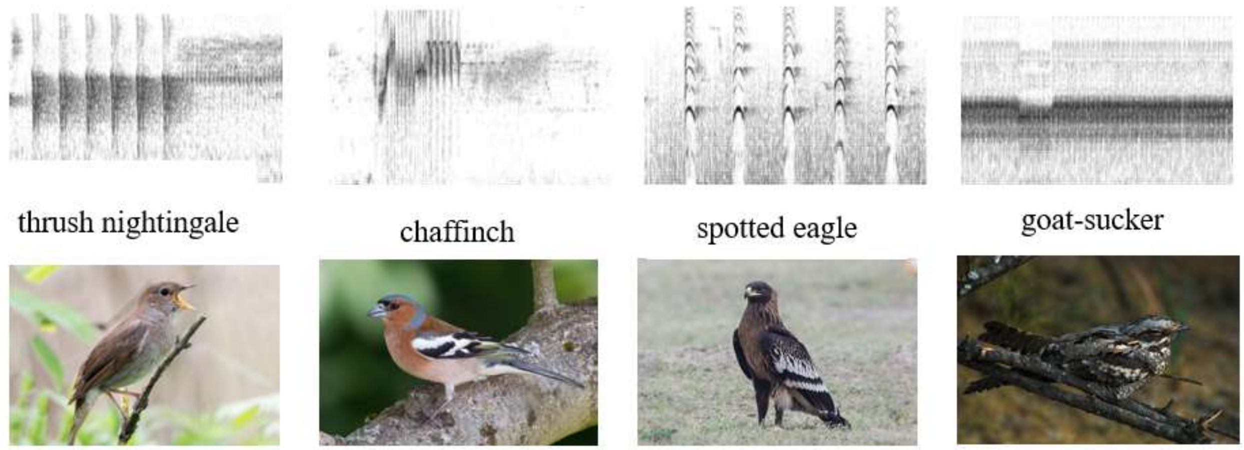

Figure 1. The spectrogram of bird voice signals

window using compression. The use of the Mel-spectral representation makes it possible to effectively describe the acoustic features of the signal, bringing them closer to the perception of the human ear and improving the quality of the features supplied to the input of the model.

The model's performance was evaluated using *precision, recall* metrics, and their geometric mean, *the f1-score*. A sample was created from the compiled dataset for testing, ensuring that the same recording did not appear in multiple samples simultaneously. Testing on a selected set of 10,000 annotations from the test data revealed that the Multi-Class Classification model achieved average precision is 0.59, recall – 0.58, and f1-score – 0.54. Meanwhile, the Multi-Label Classification model demonstrated average precision of 0.53, recall – 0.56, and f1-score – 0.52. These results are considered acceptable at this stage of the model's development.

III. The improvement of the bird sound recognition system

The above system test results prove the need to refine the application and add new functionality. The list of additional improvements, which requires additional research work. They can be described as continuing training the automated voice recognition model on more annotations and unique audio tracks for species with insufficient numbers (<200 annotations, less than 30 unique files).

To address this issue, new annotations of vocal signals have been added for bird species that previously had an insufficient number identified. Experiments were conducted to train the model using an increased number of annotations of bird vocal signals, raising the count from 250 to 400 annotations for each species. As a result of this work, a minimum threshold of 350 annotations per species was established. Plans are in place to further enhance the annotated data for those species that currently lack sufficient data to effectively train the voice recognition model, which will contribute to improved recognition accuracy.

The Multi-Label Classification architecture was employed to address the problem. The model outputs probabilities indicating the likelihood that a given sound window belongs to a specific class, thereby allowing for the prediction of multiple species occurring simultaneously. To train the automated voice recognition model, spectrograms were generated in a manner that allows multiple spectrograms to correspond to a single recording. The following experiments were conducted:

1. The minimum threshold is set at 200, and the maximum is 300 windows. The number of species that met the specified conditions was 113. The results of the experiment are shown in Table I.

Table I
The results of training the model on the number of 200-300 windows

Parameter	*precision*	*recall*	*f1-score*	*support*
micro avg	0,69	0,40	0,51	6630
macro avg	0,66	0,41	0,42	6630
weighted avg	0,65	0,40	0,42	6630
samples avg	0,42	0,42	0,42	6630

Based on the results shown in the table, it can be concluded that the minimum threshold should be set to more than 200 windows, since the recall indicator should be greater than 0.4 to ensure the accuracy of predicting the bird whose singing is present on the recording.

For example, figure 2 shows a mel-spectrogram of an audio recording with Luscinia's and Grey Warbler's singings, where the y-axis is frequency, the x-axis is time, and the signal strength is indicated by color. The spectrogram is divided into 2-second windows where the model makes predictions. Just below, the model's

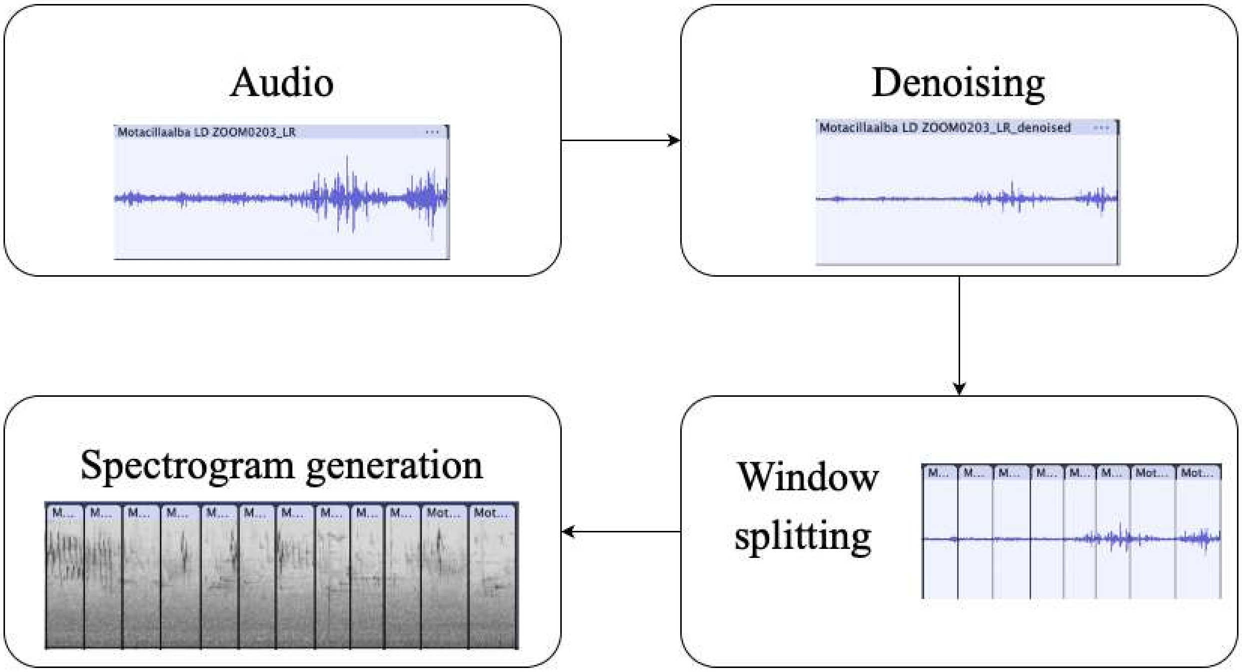

Figure 2. Processing of input audio files

predictions themselves are displayed with the model's confidence value (probability) for each 2-second window. Each bird is indicated by its own color. Everyone can see that the system indicates this audio as uscinia's song. But there is a big prabability that it can be Grey Warbler.

2. The minimum threshold set is 250, and the maximum is 350 windows. The number of species that fit this framework was 102. The following results were obtained, which are presented in Table II.

Table II
The results of training the model on the number of 250-350 windows

Parameter	*precision*	*recall*	*f1-score*	*support*
micro avg	0,90	0,74	0,81	6792
macro avg	0,88	0,74	0,79	6792
weighted avg	0,88	0,74	0,79	6792
samples avg	0,77	0,76	0,76	6792

The results improved, but in some bird species the recall index is < 0.5, which makes their singing complex. 3. The minimum threshold set is 250, and the maximum is 350 windows. The types that are poorly recognised were also subtracted (recall <0.5). The number of species that fit the noted creteria was 91. The experimental results are shown in Table III.

The decrease in the number of types in the recall indicator did not significantly affect the recognition result.

4. The minimum threshold set is 300, and the maximum is 400 windows. The number of species that

Table III
The results of training the model on the number of 250-350 windows with the recall parameter >0.5

Parameter	*precision*	*recall*	*f1-score*	*support*
micro avg	0,92	0,76	0,83	6043
macro avg	0,93	0,76	0,82	6043
weighted avg	0,92	0,76	0,82	6043
samples avg	0,77	0,77	0,77	6043

met these criteria was 85. The results are presented in Table IV.

Table IV
The results of training the model on the number of 300-400 windows

Parameter	*precision*	*recall*	*f1-score*	*support*
micro avg	0,92	0,71	0,80	6505
macro avg	0,90	0,72	0,78	6505
weighted avg	0,90	0,71	0,78	6505
samples avg	0,75	0,74	0,74	6505

As the number of windows increased, high precision values were attained; however, there are several bird species that are not well recognised in terms of recall. This indicates that these species are infrequently predicted. To address this issue, a list of birds has been created for which it is essential to increase the number of collected and processed annotations between 350 and 400: *turdusmerula, turdusphilomelos, troglodytestroglodytes, sylviaatricapilla, phylloscopuscollybita, parusmajor, luscinialus-*

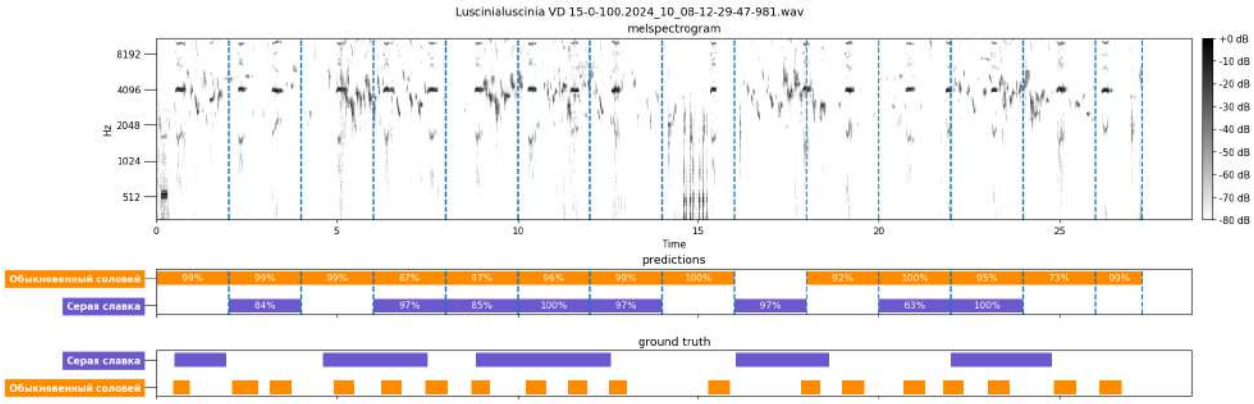

Figure 3. The recognition of Luscinia's and grey warbler's singings

cinia, fringillacoelebs, erithacusrubecula, dendrocoposmajor, cyanistescyanus , dendrocoposleucotos, corvuscornix, garrulusglandarius, pyrrhulapyrrhula, clangaclanga, scolopaxrusticola, glaucidiumpasserinum, turdusviscivorus, poecilepalustris, regulusignicapilla, dryobatesminor, ficedulahypoleuca, streptopeliaturtur, carpodacuserythrinus, alcedoatthis, picuscanus, coraciasgarrulus, leiopicusmedius, picoidestridactylus, asiootus, aquilachrysaetos, emberizacalandra, emberizahortulana, falcocolumbarius, falcotinnunculus, hieraaetuspennatus, nucifragacaryocatactes, otusscops, picusviridis, strixuralensis. This will allow everyone to train the model on more data and improve the prediction results.

Thus, to improve the efficiency of a bird voice recognition system, we are considering the following points:

1) Collect a larger and more diverse dataset of bird vocalisations, including various species, environmental conditions, and recording settings;
2) normalise audio levels and trim unnecessary segments to focus on the relevant vocalisations;
3) utilise data augmentation techniques to artificially increase the size of the training dataset by creating variations of existing recordings;
4) investigate ensemble methods that combine multiple models to improve overall accuracy;
5) implement a feedback mechanism that allows users to report misclassifications, and use this data to retrain and improve the model over time; continuously test the system on new and unseen data to evaluate its performance and identify areas for improvement; optimise the code and algorithms for faster processing times, ensuring the system can handle real-time recognition if needed.

IV. The software for the automated bird sound recognition

The experimental software designed for the automated recognition of animal (bird) vocalisations was developed on the basis of automated annotation of animal vocal signals within the collected databases and the model of recognising birds' sounds. This is the Information and Analytical Centre for Continuous Monitoring (IAC). It comprises several components, including database software, a collection of electronic datasets for training recognition models, and a website that facilitates data download and processing from various sources while generating recognition results.

To access the centre and its databases, users must register at the Information and Analytical Center. The software produces a Mel-spectrogram of the audio recording along with automated annotations of animal vocal signals based on the processed audio. These annotations can be exported into a text file. The advantages of the developed software include a user-friendly interface, a comprehensive scientific database of bird vocalisations that is continuously updated. The prototype's interface is designed to be intuitive and easy to navigate, ensuring that site visitors can easily find the information they need or accomplish their personal tasks.

On page with a database (fig. 4), each audio file is assigned a unique identification number (ID) upon download, which is displayed in the first column of the table. The subsequent columns provide additional details about the audio file, including the date (Date), recording time (RecordTime), coordinates of the bird song recording location (GPS), ambient temperature during the recording (Temperature), and the name of the recording device used (TrapName). After recognition, the name of the identified bird appears in the *BirdRecognized* column. The *SessionTime* column indicates the duration of the audio session from the time of file download to its recognition. Each audio file is also assigned a status that reflects the success of its processing.

For more information about the recognised audio file, click on (. . .). You can view a spectrogram displaying the predicted bird species and an annotation with timestamps and predicted species, as well as listen to the original audio recording.

Figure 4. The software interface of audiofiles processing

The software is fully integrated into the continuous monitoring information and analysis centre and is publicly available. Thus, the completed tasks not only produced significant scientific and technical results, but also opened up ways to further improve methods for monitoring and protecting the fauna of Belarus, contributing to the conservation and study of biodiversity in forest ecosystems. The results obtained will make it possible to spread knowledge and experience in ornithology among a wide audience and conduct further research in this direction.

V. Ontological approach for enhanced classification and monitoring of avian biodiversity

The complexity of avian biodiversity monitoring in Belarus, as highlighted in the development of the AI-powered bird sound recognition system, underscores the need for a more structured and semantically rich approach to data organisation and classification. Traditional methods of species identification, while effective to a certain extent, often struggle with the integration of diverse data types — such as audio recordings, geographic locations, and ecological contexts – leading to challenges in achieving high accuracy and scalability. To address these limitations, an ontological approach offers a promising solution by providing a formal framework for representing and interlinking knowledge about bird species, their vocalisations, and their habitats.

The recognition of bird species through vocalizations involves not only the analysis of audio signals but also the contextual understanding of where and under what conditions these vocalisations occur. For instance, the same species may exhibit variations in its calls depending on its habitat, season, or interaction with other species. Current AI models, such as CNNs, employed in this study, excel at pattern recognition but lack the ability to incorporate semantic relationships between species, their behaviours, and their environments. An ontological approach helps fill this gap by allowing knowledge to be represented in a way that machines can understand, making it easier to combine different types of data and enhancing the accuracy of classifying and monitoring species.

The OSTIS technology [9] provides a robust framework for building intelligent systems through the use of SC-code, which is a universal method of semantic representation based on semantic networks. Within OSTIS, knowledge is organised as a hierarchical system of subject domains and ontologies, allowing for the explicit description of entities, their properties, and their relationships. In the context of avian biodiversity monitoring, this approach can enhance the system's ability to reason about the relationships between bird species and their habitats, leading to more accurate identification and a deeper understanding of ecological dynamics.

To illustrate the application of OSTIS in avian biodiversity monitoring, we have integrated the description of "Bubo bubo" (Eurasian eagle-owl), a notable bird species in Belarus, with two significant geographic objects: "Belovezhskaya Pushcha" and "Berezinsky Biosphere Reserve". The description of Bubo bubo is formalized using the SC-code, following the structure provided in the study. The geographic objects are described based on the formal ontology of terrain objects outlined in next works dedicated to intelligent geoinformation systems [10].

The Bubo bubo, known as the Eurasian eagle-owl, is one of the largest owl species in the world, widely distributed across Europe, Africa, and Asia. Its description within the OSTIS framework is structured as follows:

Bubo bubo
:= [Eurasian eagle-owl]
    ∈ *English language*
:= [Puchach]
    ∈ *Belarusian language*
∈ *species*
⊂ *bubo*
    ∈ *genus*
    ⊂ *strigidae*
        ∈ *family*
        ⊂ *strigiformes*
            ∈ *order*
            ⊂ *chordata*
                ∈ *phylum*
                ⊂ *animalia*
                    ∈ *kingdom*
⇒ *habitat**:
    {• *Europe*
        ⇒ *coordinates**:
            {• [From: 48.0° N, 5.0° E]
            • [To: 65.0° N, 40.0° E]
            }
    • *Africa*
    • *Asia*
    }
⇒ *physical characteristics**:
    [Large owl with a wingspan of 1.5–2.0 meters, prominent ear tufts, and mottled brown and gray plumage]
⇒ *diet**:
    [Primarily small mammals, birds, and occasionally reptiles]
⇒ *behavior**:
    [Nocturnal, territorial, with a deep hooting call often heard at dusk or dawn]

This formalization captures the taxonomic hierarchy of Bubo bubo, its common names in English and Belarusian (sourced from the image and the Internet), its habitat range (including coordinates in Europe), and additional ecological details such as physical characteristics, diet, and behavior. The use of SC-code ensures that this knowledge is machine-readable and can be linked to other entities within the ontology.

"Belovezhskaya Pushcha" and the "Berezinsky Biosphere Reserve" are two critical habitats for avian species in Belarus, known for their rich biodiversity and conservation significance. Following the formal ontology of terrain objects proposed in [11], [12], these geographic objects are classified as areal objects within the category of vegetation cover and soils. Their descriptions in SC-code are represented below:

Belovezhskaya Pushcha
∈ *terrain object*
⊂ *areal object*
⊂ *vegetation cover and soils*
⇒ *proper name**:
    [Belovezhskaya Pushcha]
⇒ *coordinates**:
    [52.5° N, 23.8° E]
⇒ *area**:
    [1500 km²]
⇒ *vegetation type**:
    [Mixed forest with predominance of coniferous and deciduous trees]
⇒ *biodiversity significance**:
    [UNESCO World Heritage Site, habitat for rare and threatened species]
⇒ *relation of spatial-logical connections between terrain objects**:
    {• *inclusion of terrain object*
        ⇒ *requirement**:
            [Includes protected forest zones and wetland areas.]
    }

Berezinsky Biosphere Reserve
∈ *terrain object*
⊂ *areal object*
⊂ *vegetation cover and soils*
⇒ *proper name**:
    [Berezinsky Biosphere Reserve]
⇒ *coordinates**:
    [54.7° N, 28.3° E]
⇒ *area**:
    [851 km²]
⇒ *vegetation type**:
    [Coniferous forests, bogs, and meadows]
⇒ *biodiversity significance**:
    [UNESCO Biosphere Reserve, key area for migratory bird species]
⇒ *relation of spatial-logical connections between terrain objects**:
    {• *inclusion of terrain object*
        ⇒ *requirement**:
            [Includes river systems and bog ecosystems.]
    }

These descriptions provide a formal representation of the geographical objects, including their spatial properties, vegetation types, and ecological significance. The inclusion of terrain object relation indicates the presence of subentities (e.g., wetlands in Belovezhskaya Pushcha and river systems in Berezinsky Biosphere Reserve), which are critical for understanding the habitat preferences of bird species.

The integration of Bubo bubo with Belovezhskaya Pushcha and the Berezinsky Biosphere Reserve is achieved by establishing semantic relationships within the OSTIS framework. Specifically, the habitat of Bubo bubo in Europe overlaps with the coordinates of both geographic objects, allowing us to define the following relationships:

Bubo bubo
⇒ *habitat**:
  {• *Belovezhskaya Pushcha*
    ⇒ *observation frequency**:
      [Occasional, primarily in dense forest areas]
    ⇒ *vocalization characteristics**:
      [Deep, resonant hooting calls, often heard at dusk in mixed forest zones]
  • *Berezinsky Biosphere Reserve*
    ⇒ *observation frequency**:
      [Rare, observed in coniferous forests near bogs]
    ⇒ *vocalization characteristics**:
      [Lower-pitched hoots, typically in quieter, isolated areas]
  }
⇒ *relation of spatial-logical connections between species and terrain objects**:
  {• *presence in terrain object*
    ⇒ *first domain**:
      [Bubo bubo]
    ⇒ *second domain**:
      {• *Belovezhskaya Pushcha*
       • *Berezinsky Biosphere Reserve*
      }
  }

This integration links the bird species to specific habitats, detailing its observation frequency and vocalisation characteristics in each location. The presence in terrain object relation formalizes the ecological association between Bubo bubo and the two reserves, allowing the system to reason about the distribution and behavior of the species in different environments. The use of OSTIS to build an ontology for avian biodiversity monitoring offers several advantages. First, it enables the system to incorporate semantic knowledge about species and their habitats, improving the accuracy of classification by considering contextual factors such as habitat type and geographic location. For example, the distinct vocalisation characteristics of Bubo bubo in Belovezhskaya Pushcha versus Berezinsky Biosphere Reserve can be used to refine the AI model's predictions, reducing false positives in species identification. Second, the ontology facilitates the integration of diverse data sources, such as audio recordings, geographic information, and ecological metadata, into a unified knowledge base. This holistic approach supports more comprehensive analyses of avian biodiversity, such as tracking migratory patterns or assessing the impact of habitat changes on species populations. Finally, the machine-readable nature of the SC-code allows for automated reasoning and querying, enabling researchers to extract meaningful insights from the data, such as identifying key habitats for conservation prioritisation.

VI. Comparative analysis of BirdNet and AIC systems

In order to test the effectiveness of the developed system, a comparative study with the BirdNet system was carried out. *BirdNet* is an advanced machine learning system designed for the identification and classification of bird species based on their vocalisations. Developed by researchers at the Cornell Lab of Ornithology, BirdNet utilises deep learning algorithms to analyse audio recordings of bird calls and songs, enabling users to identify various bird species with high accuracy. The system can distinguish between similar-sounding species, which is often a challenge for other applications.

One of the main reasons for comparing bird voice recognition systems is to evaluate their accuracy and reliability. The comparison allows determining which system provides the best results in different conditions, such as background noise, recording quality, and variety of views. This is especially important for research, where the accuracy of identification can influence conclusions about the state of populations and ecosystems.

Comparative testing of two bird voice recognition systems can be carried out in several stages. They are defining goals and criteria, collecting data, preparing data, configuring systems, conducting testing, comparing results, statistical analysis, and interpreting statistical data.

The results of recognition of bird sound signals by BirdNet programs (desktop application, recognition model version: V2.4) and the Information and Analytical Centre for Continuous Monitoring were compared using annotated audio recordings of 232 birds singing. There are 2 audio recordings for each bird species, which are recognised by the Information and Analytical Center model. The parameters for obtaining recognition results are: minimum confidence – 80 percent, minimum bandpass frequency (Hz) – 200, maximum bandpass frequency (Hz) – 10000, sensitivity – 1. The comparison criteria were as follows:

1) the number of audio recordings with correctly predicted bird species, the number of audio recordings with unidentified or incorrectly predicted bird

Table V
The results comparing BirdNet and AIC systems

Prediction criteria	Count		Percentage (%)	
	BirdNet	AIC	BirdNet	AIC
True Positives (TP)	175	150	75	65
False Positives (FP)	57	82	25	35

species. The results are presented in the table V in quantitative and percentage terms relative to the total number of recordings;

2) The number of audio recordings with a higher maximum percentage of the prediction result (confidence threshold). The results are presented in the table V in quantitative and percentage terms relative to the number of recordings with correctly predicted bird species in both programs (150 audio recordings).

Conclusions based on the results of recognition of bird sound signals are next:

1) the Information and Analytical Centre for Continuous Monitoring correctly predicted 11 per cent fewer audio recordings than the BirdNet (out of 232 test audio recordings);
2) The Information and Analytical Center gave a higher maximum percentage of prediction results in 20 percent of audio recordings with correctly predicted bird species (150 audio recordings) compared to BirdNet.

This is a pretty high-quality result for a software environment that is currently under development. According to the test results, it is necessary to pay attention to the increase in the amount of annotated data for bird species that have not been recognised. Also, work on improving the recognition of audio recordings with poor quality will be continued.

VII. Conclusion

Bird vocalisations play a crucial role in avian behavior, ecology, and communication, making their recognition essential for fields such as ornithology, ecology, and conservation biology. This article discusses the development of an automated bird voice recognition system tailored to the unique biodiversity of Belarus, where existing applications fall short. Using artificial intelligence, particularly models of convolutional neural networks, the system aims to enhance the accuracy and efficiency of bird species identification. A comprehensive database of over 2,500 audio recordings representing 116 bird species has been created, which facilitates the training of the recognition model. The article outlines the methodology for data collection, annotation, and implementation of a user-friendly software interface that integrates various components for continuous monitoring.

After successful launch and testing, the model demonstrates the following indicators: accuracy: 0.92; macro avg: f1-score 0.5216, recall 0.5311, precision 0.92; weighted avg: f1-score 0.84, recall 0.78, precision 0.92. The results are promising, although they indicate a need for further refinement and additional data collection to improve recognition accuracy, particularly for species with insufficient annotations. Ultimately, this research contributes to the conservation of Belarusian biodiversity and promotes public participation in ornithological studies.

The creation of a comprehensive annotated dataset and the implementation of a CNN-based model lay a solid foundation for future advancements. Although the initial performance metrics are promising, ongoing efforts to expand the dataset, particularly for underrepresented species, and refine the model architecture are crucial for improving accuracy and reliability. The integration of this system within a publicly accessible information and analysis centre has the potential to transform bird monitoring efforts in Belarus, providing valuable data for conservation planning, ecological research, and public participation in ornithology. The further development and broader application of this technology promise more effective wildlife management and biodiversity conservation in the region.

In the paper, approaches to further improve the quality of ML models through the use of the ontological approach and OSTIS technology are also proposed. The integration of an ontological framework, as demonstrated with the case study of Bubo bubo (Eurasian eagle-owl) in Belovezhskaya Pushcha and Berezinsky Biosphere Reserve, enhances the system's ability to incorporate semantic relationships between species, their vocalisations, and their habitats. By formalising knowledge using SC-code within the OSTIS technology, the system can better account for contextual factors such as habitat-specific vocalisation variations, leading to a reduction in false positives and an improvement in classification accuracy beyond the current 75.6 percent. This approach also facilitates the integration of diverse data types, such as geographic and ecological metadata, into a unified knowledge base, enabling more comprehensive biodiversity analyses.

Furthermore, the ontological approach lays the groundwork for future advancements in real-time monitoring and conservation efforts. The structured representation of knowledge about species such as Bubo bubo and their habitats in Belarus supports automated reasoning and querying, which can be leveraged to identify critical areas for conservation prioritisation and to track ecological changes over time. As the project expands to include more species and data types, the ontology can be scaled to support dynamic monitoring applications, providing conservationists with actionable insights, such as immediate alerts for the presence of rare or threatened species in specific locations. This integration of semantic technology with machine learning models represents a significant step toward more accurate and ecologically informed avian biodiversity monitoring in Belarus.

Overall, as a comparative study of AIC with BirdNet shows, AIC stands out among its analogues due to its high accuracy, user-friendly design, real-time identification capabilities, and strong community engagement. These advantages make it a powerful tool for anyone interested in birdwatching, research, and conservation, helping to foster a greater appreciation for avian biodiversity.

References

[1] Sharma S., Sato K., Gautam B. P. A Methodological Literature Review of Acoustic Wildlife Monitoring Using Artificial Intelligence Tools and Techniques. Sustainability 2023, 15, № 9, pp. 1–20. Avaliable at: https://doi.org/10.3390/su15097128 (accessed 2023, Sept)

[2] Kahl St., Wood Connor M., Eibl M., Klinck H. BirdNET: A deep learning solution for avian diversity monitoring. Ecological Informatics, vol. 61, March 2021, pp. 1–10. Avaliable at: https: //doi.org/10.1016/j.ecoinf.2021.101236 (accessed 2023, May).

[3] Hetsevich Yu .S., Dydo O. V., Bialiauski D. A. Technologies of automatic speech processing and analysis using artificial intelligence. Doklady II Foruma IT-Akademgrada «Iskusstvennyj intellekt v Belarusi» [Reports of the II IT Academy Forum "Artificial Intelligence in Belarus", Minsk, October 12–13, 2023], Minsk, UIIP of NASB, 2023, pp. 71–78 104 (In Bel).

[4] Miao Z., Gaynor K. M., Wang J., Liu Z., Muellerklein O., Norouzzadeh M.S. Insights and approaches using deep learning to classify wildlife. Scientific Reports, 2019, vol. 9, № 1, p. 10. Avaliable at: https://doi.org/10.1038/s41598-019-44565-w (accessed 2023, Dec).

[5] ZianoukaYa., Bialiauski D., Kajharodava L., Chachlou V., Hetsevich Yu., Zahariev V., Zhaksylyk K. Developing Birds Sound Recognition System Using an Ontological Approach. Open Semantic Technologies for Intelligent Systems, Minsk, Belarusian State University of Informatics and Radioelectronics, 2023, issue 7, pp. 165–170.

[6] Hiecevič, J.S., Zianoŭka J.S., Trafimaŭ A.S., Bakunovič A.A., Latyševič D.I., Drahun A.JA., Sliesarava M.M., Tukaj M.S., Komplieks srodkaŭ realizacyi zadač štučnaha 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)

[7] Mingxing Tan, Quoc V. Le EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv, 1905.11946v5, 2019. Available at: https://arxiv.org/abs/1905.11946 (accessed 15 February 2024).

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

[9] Tehnologija kompleksnoj podderzhki zhiznennogo cikla semanticheski sovmestimyh intellektual'nyh komp'juternyh sistem novogo pokolenija / pod obshh. red. V. V. Golenkova, Minsk, Bestprint, 2023, pp. 1064.

[10] Samodumkin S. Designing Intelligent Systems with Integrated Spatially Referenced Data. Open Semantic Technologies for Intelligent Systems (OSTIS), BSUIR, Minsk, 2024, Iss. 8, pp. 195–212.

[11] Samodumkin S. Next-generation intelligent geoinformation systems. Open Semantic Technologies for Intelligent Systems (OSTIS), BSUIR, Minsk, 2022, Iss. 6, pp. 411–416.

[12] Samodumkin S. Semantic Technology of Intellectual Geoinformation Systems Development. Open Semantic Technologies for Intelligent Systems (OSTIS-2019), BSUIR, Minsk, 2019, pp. 231 - 236.

РАЗРАБОТКА СИСТЕМЫ РАСПОЗНАВАНИЯ ГОЛОСОВ ПТИЦ ДЛЯ МОНИТОРИНГА БИОРАЗНООБРАЗИЯ ФАУНЫ БЕЛАРУСИ

Зеновко Е., Дыдо О., Дыдо А., Жалова Д., Гецевич Ю., Захарьев В., Крищенович В.

В статье рассматривается разработка автоматизированной системы распознавания голосов птиц, адаптированной к уникальному биологическому разнообразию Беларуси. Система, разработанная на основе технологий искусственного интеллекта, в частности глубокого обучения моделей на основе сверточных нейронных сетей, направлена на повышение точности и эффективности идентификации вокализаций видов птиц. Созданная обширная база данных из более чем 2500 аудиозаписей содержит вокализации 116 видов птиц, что содействует обучению модели распознавания на разнообразных голосах птиц.

В статье описывается методология сбора данных, аннотирование и внедрение удобного программного интерфейса, который объединяет различные компоненты для непрерывного мониторинга. После успешного запуска и тестирования модель демонстрирует следующие показатели: точность: 0,92; макро среднее: f1-мера 0,5216, полнота 0,5311, точность 0,92; средневзвешенное значение: f1-мера 0,84, полнота 0,78, точность 0,92. Результаты многообещающие, хотя и указывают на необходимость дальнейшего уточнения и сбора дополнительных данных для повышения точности распознавания, особенно для редких видов с недостаточным количеством аннотаций. Исследование способствует сохранению биоразнообразия Беларуси и содействует участию общественности в орнитологических исследованиях.

Интеграция алгоритмов распознавания в общедоступный информационно-аналитический центр (ИАЦ) содействует объединению усилий по мониторингу птиц в Беларуси, предоставляя ценные данные для планирования охраны природы, экологических исследований и участия общественности в орнитологии.

Как показывает сравнительное исследование информационно-аналитического центра с автоматической системой распознавания голосовых сигналов BirdNet, ИАЦ выделяется среди своих аналогов благодаря высокой точности, удобному дизайну, возможностям идентификации в режиме реального времени и активному участию сообщества. Эти преимущества делают его эффективным инструментом для всех, кто стремится лучше понимать биоразнообразие птиц.

В статье также предлагаются подходы к дальнейшему повышению качества ML-моделей за счет использования онтологического подхода и технологии OSTIS. Интеграция онтологической структуры, как было продемонстрировано на примере Бубо-бубо (евразийского филина) в Беловежской пуще и Березинском биосферном заповеднике, повышает способность системы учитывать семантические связи между видами, их вокализациями и местами обитания. Формализуя знания с помощью SC-кода в рамках технологии OSTIS, система может лучше учитывать контекстуальные факторы, такие как вариации вокализации в зависимости от среды обитания, что приводит к снижению числа ложных срабатываний и повышению точности классификации по сравнению с текущими 75,6 процентами.

Received 23.03.2025