

ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ, РЕЧИ, ТЕКСТА И РАСПОЗНАВАНИЕ ОБРАЗОВ

SIGNAL, IMAGE, SPEECH, TEXT PROCESSING AND PATTERN RECOGNITION



УДК 598.2:591.582.2:004.934
DOI: 10.37661/1816-0301-2024-21-4-58-71

Арыгінальны артыкул
Original Article

Аўтаматызацыя аналізу галасавых сігналаў птушак

Ю. С. Гецэвіч, Я. С. Зяноўка[✉], А. А. Бакуновіч, Д. А. Жалава, Т. Г. Шагава

*Аб'яднаны інстытут праблем інфарматыкі
Нацыянальнай акадэміі навук Беларусі,
вул. Сурганава, 6, Мінск, 220012, Беларусь
[✉]E-mail: evgeniakacan@gmail.com*

Анотацыя

Мэты. Мэтай працы з'яўляецца стварэнне эксперыментальнага праграмнага забеспячэння аўтаматызаванага распазнавання галасавых сігналаў, якое валодае магчымасцямі доўгачасовага кругласучаснага і кругласезоннага маніторынгу відавой разнастайнасці жывёл у абраных месцапражываннях і экасістэмах. Метады. У працы выкарыстоўваюцца метады глыбокага машыннага навучання скруткавых нейронных сетак, навучаных на падставе мел-спектраграм вакалізацый птушак, якія пабудаваны з дапамогай хуткага пераўтварэння Фур'е.

Вынікі. Апісаны працэс, метады і падыходы да трэніроўкі мадэлі глыбокага машыннага навучання для перспектыўнай сістэмы пасіўнага акустычнага маніторынгу папуляцый птушак на тэрыторыі Беларусі, а таксама ў ходзе тэсціравання прататыпа праграмнага забеспячэння выяўлены складанасці і дасягнуты вынікі.

Заклучэнне. Прадстаўлены працоўны прататып праграмнага забеспячэння аўтаматызаванага распазнавання галасавых сігналаў птушак. Ён ажыццяўляе аналіз акустычных запісаў галасоў птушак з выдачай імавернаснай ацэнкі відавой прыналежнасці для прысутных на запісах вакалізацый жывёльнага паходжання. Праграмнае забеспячэнне накіравана на павышаную эфектыўнасць маніторынгу птушак, што забяспечвае выкананне прыродаахоўчых і даследчых мерапрыемстваў на аснове дакладных і актуальных даных відавочнага распаўсюджвання.

Ключавыя словы: аўтаматызаванае распазнаванне галасавых сігналаў, скруткавыя нейронныя сеткі, глыбокае машыннае навучанне, мел-спектраграма, анатаванне галасавых сігналаў

Падзякі. Даследаванне ажыццяўляецца ў рамках тэхнічнага задання на выкананне НДП па дамове № 221/35 ад 06.09.2021 г. па тэме «Распрацоўка тэхналогіі аўтаматызаванага распазнавання галасавых сігналаў жывёл для ажыццяўлення аўтаномнага бесперапыннага маніторынгу рэдкіх відаў, відаў з пагрозамі і індыкатарных відаў і стану біразнастайнасці ў лясных экасістэмах (у частцы тэхнічнай рэалізацыі)» сумесна з НПЦ НАН Беларусі па біярэсурсах.

Для цытавання. Аўтаматызацыя аналізу галасавых сігналаў птушак / Ю. С. Геце́віч, Я. С. Зяноўка, А. А. Бакуновіч [і інш.] // Інформатика. – 2024. – Т. 21, № 4. – С. 58–71. – DOI: 10.37661/1816-0301-2024-21-4-58-71.

Канфлікт інтарэсаў. Аўтары заяўляюць аб адсутнасці канфлікту інтарэсаў.

Паступіла ў рэдакцыю | Received 05.08.2024

Падпісана ў друк | Accepted 27.09.2024

Апублікавана | Published 30.12.2024

Automation of bird voice signal analysis

Yuras S. Hetsevich, Yauheniya S. Zianouka[✉], Andrei A. Bakunovich, Darja A. Zhalava, Tatsiana G. Shagava

*The United Institute of Informatics Problems
of the National Academy of Sciences of Belarus,
st. Surganova, 6, Minsk, 220012, Belarus
✉E-mail: evgeniakacan@gmail.com*

Abstract

Objectives. The purpose of the work is to create an experimental software for automated recognition of voice signals, which has the capabilities of long-term round-the-clock and round-the-season monitoring of animal species diversity in selected habitats and ecosystems.

Methods. The work uses methods of deep machine learning of convolutional neural networks trained on mel-spectrograms of bird vocalizations, which are built using fast Fourier transform.

Results. The process, methods and approaches to training a deep machine learning model for a system of passive acoustic monitoring of bird populations in Belarus are described, as well as the difficulties identified during testing of the software prototype and the results that were achieved.

Conclusion. A working prototype of the software for automatic recognition of animal (bird) voice signals is presented. It performs the analysis of acoustic recordings of bird voices with the issue of probabilistic assessment of species belonging to animal vocalizations present in the recordings. The software is aimed at increasing the efficiency of bird monitoring, which ensures the implementation of conservation and research activities based on accurate and up-to-date data on species distribution.

Keywords: automatic recognition of voice signals, convolutional neural networks, deep machine learning, mel-spectrogram, annotation of voice signals

Acknowledgements. The research is carried out within the research work under contract No. 221/35 dated September 6, 2021 on the title "Development of technology for automated recognition of animal voice signals for autonomous continuous monitoring of rare, threatened and indicator species and the state of biodiversity in forest ecosystems (in terms of technical implementation)" jointly with the Scientific and Practical Center of the National Academy of Sciences of Belarus for Bioresources.

For citation. Hetsevich Y. S., Zianouka Ya. S., Bakunovich A. A., Zhalava D. A., Shagava T. G. *Automation of bird voice signal analysis*. Informatika [Informatics], 2024, vol. 21, no. 4, pp. 58–71 (In Bel.). DOI: 10.37661/1816-0301-2024-21-4-58-71.

Conflict of interest. The authors declare of no conflict of interest.

Уводзіны. У прыродзе існуе шмат тыпаў галасавых сігналаў, якія ажыццяўляюць разнастайныя функцыі. У асобную групу падобных сігналаў адносяць галасавыя спевы птушак – гэта больш меладычны тып галасоў, звычайна даўжэй і складаней, чым кліч, рык, шыпенне і інш. Распазнаванне галасоў птушак мае некалькі важных аспектаў. Па-першае, маніторынг біразнастайнасці для адсочвання папуляцыі, ацэнкі іх колькасці і стану здароўя экасістэмы.

Па-другое, анализ сацыяльных структур, тэрытарыяльных паводзін і рэпрадуктыўных стратэгіі птушак. Гэта можа дапамагчы птушкам у адаптацыі да змен навакольнага асяроддзя. Па-трэцяе, кантроль якасці асяроддзя пражывання і выяўлення парушэнняў (выруб лясоў або забруджванне), якія могуць паўплываць на птушыныя папуляцыі. Пералічаныя фактары падкрэсліваюць важнасць і значэнне выяўлення галасоў птушак у розных кантэкстах [1].

З-за мноства варыяцый гукавых сігналаў розных відаў птушак узнікае праблема іх распазнавання. Зварот да спецыялістаў-арнітолагаў патрабуе дадатковага часу на ручную апрацоўку асобных сігналаў, што значна ўскладняе працэс аўтаматызацыі вызначэння асобнага віда. Гэта сведчыць аб неабходнасці распрацоўкі сістэм аўтаматызаванага распазнавання галасоў птушак для стварэння мадэляў, якія могуць аналізаваць вялікія аб'ёмы даных [2]. На сённяшні дзень вядомы прыкладанні: *Aubio*, *prodigy*, *BirdNET*, *Merlin Bird ID by Cornell Lab*, *Song Sleuth*, *Warblr*, *Chirpomatic*, якія разглядаюць галасавыя сігналы птушак на розных тэрыторыях свету. Актуальнасць стварэння сістэмы распазнавання галасоў птушак на тэрыторыі Беларусі абумоўлена тым, што ўсе існуючыя прыкладанні па распазнаванню не ахопліваюць усю біразнастайнасць птушак Беларусі, толькі некаторыя з іх здольны генэраваць гукавыя сігналы відаў Еўропы. Акрамя таго, аўтаматычнае распазнаванне галасавых сігналаў з'яўляецца асабліва вострай праблемай у краінах, якія выпрабавваюць недахоп магчымасцяў для эфектыўнага маніторынгу і аховы фаўны. Паспяховае развіццё такіх сістэм можа мець шырокае практычнае прымяненне, уключаючы больш эфектыўныя сродкі маніторынгу папуляцый птушак, ацэнкі іх экалагічнага стану і захавання біразнастайнасці [3].

Праграмнае забеспячэнне для вызначэння відаў птушак па галасавых сігналах грунтуецца на мадэлях матэматычных вылічэнняў (часцей за ўсё скруткавых нейронных сетках), якія пры недастатковай навучанасці могуць зрабіць вылічальную памылку, што прыводзіць да няправільнага вызначэння неабходнага біялагічнага віду [4, 5]. Такім чынам, у рамках рэалізацыі праекта «Распрацоўка тэхналогіі аўтаматызаванага распазнавання галасавых сігналаў жывёл для ажыццяўлення аўтаномнага бесперапыннага маніторынгу рэдкіх відаў, відаў з пагрозамі і індыкатарных відаў і стану біразнастайнасці ў лясных экасістэмах (у частцы тэхнічнай рэалізацыі)» пастаўлены наступныя задачы [6]:

- падрыхтоўка метадычных асноў для збору, анатавання і распазнавання галасавых сігналаў жывёл на тэрыторыі Беларусі (у частцы тэхнічнай рэалізацыі);

- арганізацыя масіваў даных наяўных электронных карпусоў галасавых сігналаў жывёл з даступных крыніц;

- распрацоўка структурнай схемы аўтаматызаванага распазнавання галасавых сігналаў жывёл для ажыццяўлення аўтаномнага бесперапыннага маніторынгу рэдкіх відаў, відаў з пагрозамі і індыкатарных відаў;

- стварэнне інфармацыйна-аналітычнага цэнтра аўтаномнага бесперапыннага маніторынгу рэдкіх відаў, відаў з пагрозамі і індыкатарных відаў жывёл (у частцы тэхнічнай рэалізацыі);

- генерацыя аўтаматызаванай разметкі галасавых сігналаў жывёл у сабраных базах даных.

Іх выкананне ажыццяўляецца лабараторыяй распазнавання і сінтэзу маўлення Аб'яднанага інстытута праблем інфарматыкі НАН Беларусі (<https://ssrlab.by/>).

Камп'ютарныя метады класіфікацыі гукавых сігналаў. Класіфікацыя гукавых сігналаў птушак (жывёл) з'яўляецца важнай задачай у заалогіі і біяакустыцы. Для яе вырашэння прымяняюцца розныя метады:

1. Акустычны анализ – выкарыстоўвае спектраграмныя метады для візуалізацыі гукавых сігналаў. Спектраграма дазваляе ўбачыць частату, амплітуду і часовыя характарыстыкі гуку. Анализ асноўных частот і іх гармонік дапамагае адрозніваць сігналы розных відаў.

2. Машыннае навучанне – уключае метады класіфікацыі і глыбокага навучання. Метады класіфікацыі ўяўляюць алгарытмы машыннага навучання, такія як k-бліжэйшых суседзяў (k-nearest neighbors, KNN), дрэвы рашэнняў, выпадковыя лясы і нейронныя сеткі для класіфікацыі гукаў на аснове характарыстык, вынятых з гукавых сігналаў. Глыбокае навучанне – больш прасунуты метад, у тым ліку і скруткавыя нейронныя сеткі, можа выкарыстоўвацца для аўтаматычнай класіфікацыі гукавых сігналаў, здабываючы складаныя патэрны з сырых гукавых даных.

3. Статыстычныя метады – прымяняюцца для аналізу і інтэрпрэтацыі акустычных даных, для групування гукаў па падабенстве – тэсты на значнасць і кластарны аналіз.

4. Стварэнне баз даных акустычных сігналаў розных відаў, якія выкарыстоўваюцца для навучання і тэсціравання класіфікатараў.

5. Параўнальны аналіз – ажыццяўляе супастаўленне гукавых сігналаў розных відаў па розных параметрах, такіх як даўжыня, рытм, танальнасць і мадуляцыя.

6. Кантэкстуальны аналіз – улічвае кантэкст, які выдае гукі, напрыклад паводзіны жывёл, для лепшага разумення значэння гукаў.

7. Інтэрдyscyплінарныя падыходы – выкарыстоўваюць метады з іншых абласцей, такіх як псіхалогія і нейрабіялогія, для разумення кагнітыўных працэсаў, звязаных са стварэннем і ўспрыманням гукавых сігналаў.

У рамках праведзенага даследавання абраны метады глыбокага навучання, якія пераўзыходзяць традыцыйныя метады класіфікацыі гукавых сігналаў жывёл па некалькіх прычынах [7]. Глыбокія нейронныя сеткі, асабліва скруткавыя, здольны аўтаматычна здабываць рэпрэзентатыўныя прыкметы з гукавых сігналаў праз мел-спектраграмы без неабходнасці ручнога рэдагавання. Гэта дазваляе пазбягаць страт інфармацыі, якія могуць паўстаць пры папярэдняй апрацоўцы і выбары прыкмет. Нейронныя сеткі здольны ахопліваць больш тонкія і складаныя дэталі гукавых сігналаў, уключаючы змены ў частаце, амплітудзе і рытме, што важна пры класіфікацыі. Глыбокае навучанне выдатна спраўляецца з вялікімі аб'ёмамі даных, можа ўлічваць складаныя патэрны і ўзаемасувязі ў даных. Падобныя мадэлі больш устойлівыя да шуму і скажэнняў, паколькі яны навучаюцца на разнастайных даных з рознымі ўзроўнямі фону і шумоў. Скруткавыя нейронныя сеткі лёгка адаптаваны для вырашэння разнастайных задач, менавіта класіфікацыі, сегментацыі і нават анатацыі даных [8]. Напрыклад, адну і тую ж архітэктuru можна наладзіць пад розныя віды гукавых сігналаў, змяніўшы даныя навучання. Гэтыя перавагі робяць глыбокае навучанне магутным інструментам для аналізу і класіфікацыі гукавых сігналаў птушак, дазваляючы даследчыкам здабываць больш дакладныя і надзейныя вынікі ў параўнанні з традыцыйнымі метадамі.

Масівы даных наяўных электронных карпусоў галасавых сігналаў птушак (жывёл). Для паспяховай рэалізацыі стварэння аўтаматызаванай сістэмы распазнавання галасоў птушак неабходна анатаваная база даных аўдыязапісаў галасавых сігналаў птушак. База даных уяўляе сховішча гукавых запісаў і тэкставых файлаў з анатацыямі, якія выкарыстоўваюцца для навучання мадэлі праекту [9]. У цяперашні час вядомы разнастайныя прыкладанні і вэб-сайты з адкрытым доступам, якія змяшчаюць вялікія аб'ёмы гукавых даных па розных таксонах жывёл з розных рэгіёнаў свету [10]. Напрыклад, *Avibase*, *eBird*, *iNaturalist*, *NIPS4B*, *Freefield*, *Chernobyl Exclusion Zone Dataset* і інш. У рамках праведзенага даследавання выбар базы даных прадугледжвае адбор гукавых файлаў відаў, якія насяляюць рэгіёны Беларусі, ці карыстанне пры пошуку меткамі рэгіёнаў, фаўна якіх перакрывавацца з беларускай.

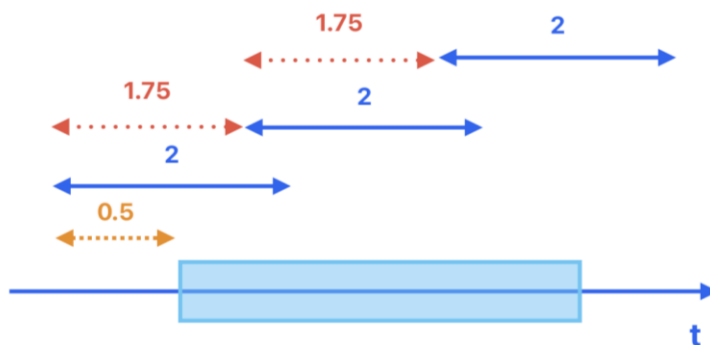
Па актуальных даных, арнітафаўна Беларусі ўключае 342 віды [11]. Прататып сістэмы галасавых сігналаў птушак павінен ажыццяўляць дакладную відавую класіфікацыю сярод 116 відаў, якія характарызуюцца шырокім распаўсюджваннем на тэрыторыі Беларусі і выкарыстаннем ярка выяўленых вакалізацый у ходзе пасіўнага акустычнага маніторынгу на лакацыях. Самі вакалізацыі для навучання і тэсціравання мадэлі праграмага забеспячэння сабраны з уласнага масіву гукавых сігналаў, а таксама з корпуса даных *Xeno-Canto*¹. Гэта адна з найбуйнейшых крыніц гукавых даных галасавых сігналаў птушак, сабраных з усяго свету. Сайт мае API-эндпойнты, якія магчыма выкарыстоўваць для аўтаматычнага пошуку, спампавання даных па навуковай ці распаўсюджанай назве віда альбо сямейства, тэгах рэгіёну, тыпах гуку, краіне і г. д. На май 2024 г. крыніца змяшчае 857 584 запісы галасоў. Дабаўленне запісаў даступна для карыстальнікаў пасля рэгістрацыі, ёсць магчымасць выкарыстоўваць дадзены корпус для загрузкі асабіста сабраных запісаў разам з запісамі іншых. Кожны файл анатаваны прафесійным арнітолагам адзнакамі пачатку і канца вакалізацыі віду на запісы, а таксама фонавымі вакалізацыямі іншых відаў і іншых шумоў.

¹Xeno-canto – sharing wildlife sounds from around the world. – 2024. – URL: <https://xeno-canto.org/> (date of access: 01.04.2024).

Такім чынам, масіў даных электронных карпусоў галасавых сігналаў (датасэт), – гэта кампанент інфармацыйна-аналітычнага цэнтра, на падставе якога выконваецца навучанне мадэлі распазнавання. Ён створаны на аснове ўзораў аб'яднаных карпусоў запісаў клічаў і спеваў птушак з адкрытых крыніц і даных, падрыхтаваных супрацоўнікамі НПЦ НАН Беларусі па біярэсурсах. Пры падрыхтоўцы даных для навучання быў створаны файл .csv з інфармацыяй пра анатацыі ад цэнтра з пазнакамі часу пачатку і канца анатацый, а таксама са шляхамі да адпаведных аўдыязапісаў і спектраграм, захаваных у фармаце .npy.

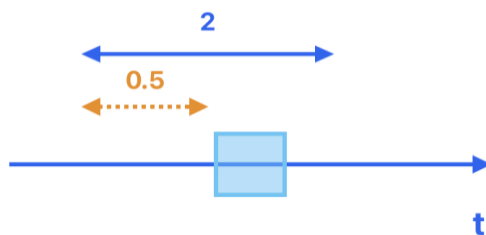
На першапачатковым этапе распрацоўкі мадэлі і алгарытму ігнараваліся фонавыя гукі і шумы [12]. Так як архітэктурой мадэлі прадугледжаны фіксаваны памер уваходнага тэнзару, то для навучання спектраграмы аналізаваліся вокнамі, памер аднаго акна роўны 2 с. З-за рознай працягласці анатацый рэалізаваны алгарытм падрыхтоўкі адрэзкаў працягласцю, роўнай даўжыні акна. Алгарытм прапануе два варыянты апрацоўкі аўдыяфайлаў:

Варыянт 1. Калі працягласць анатацыі не менш за палову даўжыні акна, то пазнака часу пачатку першай анатацыі выбіралася з выпадковым зрухам ад 0 да 0,5 с улева. Пасля выкарыстоўваўся зрух у 1,75 с для генерацыі большай колькасці анатацый на падставе старой. На мал. 1 прадстаўлена графічнае адлюстраванне варыянта 1.



Мал. 1. Апісанне алгарытму першага варыянта апрацоўкі файлаў
Fig. 1. Description of the algorithm for the first file processing option

Варыянт 2. У выпадку, калі працягласць анатацый менш за палову даўжыні акна, то адзіночная маленькая анатацыя ўся ўключалася ў дзвюхсекундную анатацыю з выпадковым зрухам улева даўжынёй да 0,5 с. Мал. 2 адлюстроўвае дадзены выпадак.



Мал. 2. Апісанне алгарытму другога варыянта апрацоўкі файлаў
Fig. 2. Description of the algorithm of the second file processing option

Такім чынам апрацавана больш за 2500 аўдыязапісаў і атрымана каля 42 000 анатаваных вакалізацый для 116 відаў птушак. Набор даных уключае запісы рознай якасці, а таксама запісы з рознай колькасцю сумесна вакалізуючых відаў. Варта адзначыць, што матэрыял не сбалансаваны. Для далейшага паляпшэння мадэлі быў створаны клас *background*, які ўтрымлівае розныя тыпы шумоў і цішыню. Для гэтага браліся неанатаваныя дзвюхсекундныя адрэзкі аўдыяфайлаў або з меткамі шумоў без перасячэння з анатацыямі птушак. Для навучання мадэлі файлы з анатацыямі відаў птушак і фона былі аб'яднаны ў адзіны датафрэйм.

Праграмнае забеспячэнне для навучання мадэлі распазнавання. Існуе некалькі падыходаў да пабудовы сістэм распазнавання галасавых сігналаў [13]:

1. Мадэлі распазнавання на аснове спектраграм, у рамках якіх аўдыясігнал пераводзіцца ў спектраграму – візуальную рэпрэзентацыю частот сігнала на працягу часу. Выкарыстоўваюцца скруткавыя нейронныя сеткі (Convolutional Neural Network, CNN), такіх від нейронных сетак больш прывабны для задач распазнавання выяў.

2. Мадэлі на аснове амплітуднай кампаненты сігнала (сігнала без аналізу частотных характарыстык), у якіх выкарыстоўваюцца рэкурэнтныя сеткі.

3. Мадэлі на аснове сінтэза спецыфічных карысных характарыстык сігналаў, такіх як мел-кепстральныя каэфіцыенты (MFCCs), кадыраванне з лінейным прадказальнікам (LPC), гаматонавыя фільтры (gammatone filterbanks).

4. Гібрыдныя мадэлі – уключаюць некалькі тыпаў мадэляў для сінтэзу мадэлі распазнавання, каб атрымаць лепшую якасць распазнавання.

5. Трансфернае навучанне (transfer learning) – прапануе навучанне на аснове папярэдне навучанай мадэлі на іншых даных, дзе вагі ўжо ўстаноўлены. Мадэль пры гэтым данавучаецца на наяўных даных аўдыязапісаў.

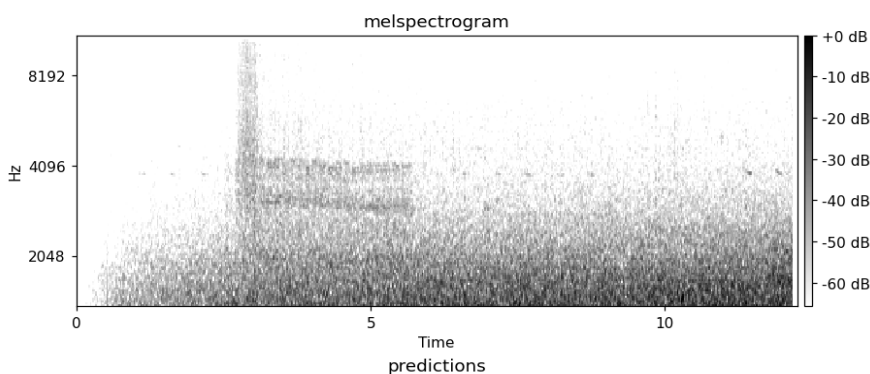
У працы абраны падыход пабудовы мадэляў распазнавання на аснове мел-спектраграм, пры якім атрымліваецца найлепшая якасць генерацыі гукавых сігналаў. Спектраграма – гэта візуалізацыя мноства частот на адным графіку, дзе на восі X адлюстроўваецца час, на восі Y – частата, а яркасць або колер адлюстроўваюць інтэнсіўнасць гуку або энергію сігнала на канкрэтнай частаце ў пэўны момант часу. Асноўнае адрозненне мел-шкалы ад звычайнай частотнай шкалы заключаецца ў тым, што адлегласці паміж кропкамі на мел-шкале нелінейна павялічваюцца з павелічэннем частаты. Гэта робіць мел-спектраграму больш набліжанай да здольнасці чалавечага слыху ўспрымаць гукі рознай вышыні, улічваючы асаблівасці чалавечага слуху (адсейваючы частоты гукаў, якія чалавек не чуе). Так як птушкі пяюць высокімі частотамі, прымяняецца фільтр верхніх частот, каб убраць непатрэбны шум (пакінуць частоты мінімальнай велічыні ў 200 Гц). Для генерацыі мел-спектраграмы выкарыстана бібліятэка для мовы праграмавання Python – librosa. Менавіта мадыфікаваная спектраграма і будзе падавацца на ўваход мадэляў распазнавання.

Для генерацыі мел-спектраграмы была выкарыстана бібліятэка на Python librosa:

```

signal, sr = librosa.load(fp, sr=self.sr, duration=self.duration, mono=self.mono)
S_ms = librosa.feature.melspectrogram(y=signal, sr=sr, n_fft=self.n_fft,
hop_length=self.hop_length, n_mels=self.n_mels, fmin=self.fmin, htk=self.is_htk)
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.
    
```

Мел-спектраграма голаса птушкі *Columba Palumbus* (вяхір) з адзначэннем фільтра верхніх частот і вылічэннем мел-кепстральных каэфіцыентаў адлюстравана на мал. 3.



Мал. 3. Мел-спектраграма аўдыясігнала птушкі *Columba Palumbus* (вяхір)

Fig. 3. Mel-spectrogram of the *Columba Palumbus* bird (vjahir)

На ўваход сістэмы паступае аўдыяфайл, а на выхадзе атрымліваецца паслядоўнасць лэйблаў (назвы класаў для птушак) [14]. Кожны n -ны лэйбл – гэта вынік распазнавання галаса птушкі n -га фрэйма даўжынёй $duration\ ms$ на прамежку часа.

$[(n-1)*lag, (n-1)*lag + duration]$, дзе lag – шкала для здвіга акна.

Распазнаванне віда птушкі адбываецца праз падрыхтаваны спіс мел-спектраграм для кожнага віду (класа). Мадэль вылучае спецыфічныя характарыстыкі апрацаваных даных (выяў), падае іх на ўваход глыбокай нейроннай сетцы і выдае набор верагоднасцяў, якія адпавядаюць асобнаму класу [15].

У даследаванні выкарыстоўваецца архітэктара скруткавых нейронных сетак *EfficientNetB3*, а таксама яшчэ тры слоя (*Flatten*, *Dropout*, *Dense* з функцыяй актывацыі *Softmax* у якасці выхада) для пабудовы мадэлі CNN. *EfficientNet* – сучасная распрацоўка скруткавай нейроннай сеткі ад Google Brain [16]. Асноўная ідэя *EfficientNet* заключаецца ў камбінаваным маштабаванні памеру скруткавых нейронных сетак. Напрыклад, можна павялічыць скруткавую сетку на аснове шырыні слаёў, глыбіні слаёў, памеры ўваходнага малюнка або камбінацыі ўсіх гэтых параметраў. Такім чынам, мадэль распазнавання для кожнага з акон была пабудавана на аснове *EfficientNetB3* з аптымізатарам *Adam* (*Adam optimizer*), функцыяй страт катэгарыяльнай перакрываемай энтрапіі (*categorical cross-entropy loss function*) і збалансаванымі вагамі класаў.

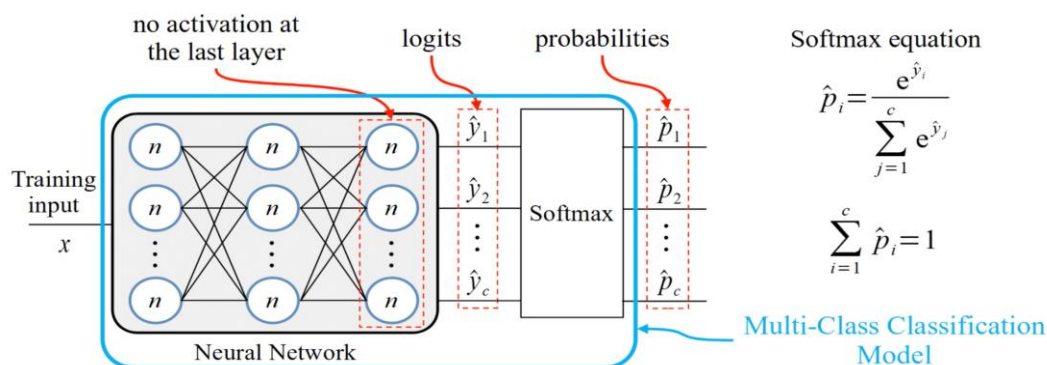
Для таго каб прыстасаваць выходныя даныя *EfficientNetB3* да задачы класіфікацыі, да сеткі дадаюцца наступныя слаі:

Flatten – пераўтварае шматмерны тэнзар, які атрымліваецца з апошняга слоя *EfficientNetB3*, у аднамерны вектар. Неабходнасць гэтага дзеяння ў тым, каб далучыць вектар да наступных поўназлучных слаёў (*Dense*), якія працуюць з аднамернымі ўваходнымі данымі.

Dropout – прадухіляе перанавучанне мадэлі. Ён выпадковым чынам абнуляе пэўны працэнт нейронаў падчас навучання (у гэтым даследаванні 40 %), што дапамагае мадэлі не канцэнтравцца на пэўных характарыстыках даных і паляпшае яе агульную здольнасць да абагульнення.

Dense – гэта поўназлучны слой, які і робіць прадказанні аб прыналежнасці даных да класаў.

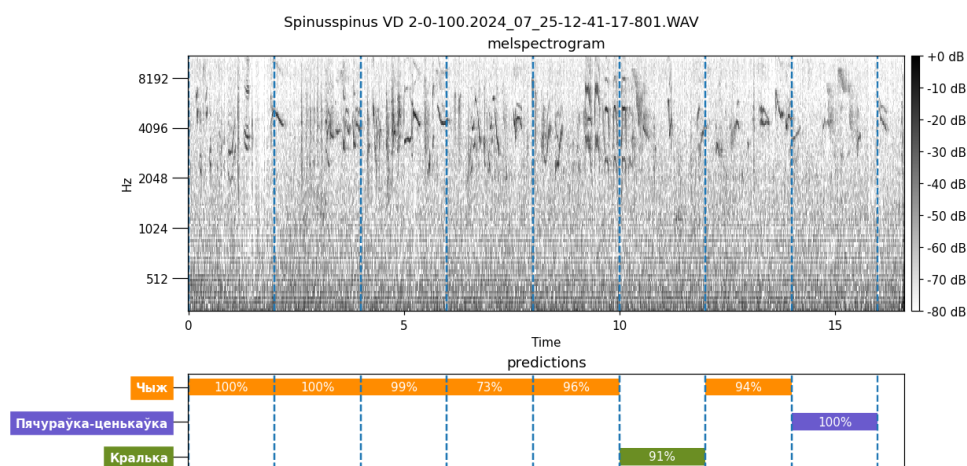
У выніку праведзеных прац распрацавана мадэль, якая вырашае задачу *Multi-Class Classification* [17]. Яе сутнасць палягае ў тым, што аднаму акну аўдыясігналу ставіцца ў адпаведнасць прадказанне аднаго віда. Для такіх мадэляў у якасці функцыі актывацыі выкарыстоўваецца актыватар *Softmax*. На выхадзе працы мадэлі атрымліваецца размеркаванне імавернасцей такім чынам, што прадказанаму віду прыпісваецца значэнне імавернасці не рэальнае, а зрушанае ў бок да класа з найбольшай вагай функцыі актывацыі *Softmax*. Таму мадэль прадказвае адзін від на абраным акне распазнавання (мал. 4).



Мал. 4. Архітэктара мадэлі Multi-Class Classification з функцыяй актывацыі Softmax

Fig. 4. Architecture of the Multi-Class Classification with the function of Softmax

Актыватар *Softmax* размяркоўвае імавернасці такім чынам, каб на больш верагодным класе імавернасць была больш і сума па ўсім класам была роўная аднаму. Таму звычайна на больш верагодным класе імавернасць значна большая, чым на астатніх. На мал. 5 прапанаваны выгляд працы мадэлі з функцыяй актывацыі *Softmax*.



Мал. 5. Спектраграма анатавання мадэлі з адлюстраваннем верагоднасцей
 Fig. 5. Spectrogram of model annotation with probability display

Тэсціраванне працы мадэлі пасля трэніроўкі па 116 відах з дадатковым класам *Unknown* праводзілася на аснове метрык, якія выкарыстоўваюцца для ацэнкі эфектыўнасці мадэлі класіфікацыі. Кожная з іх ацэньвае прадуктыўнасць мадэлі адпаведна розным крытэрыям:

дакладнасць (precision) – гэта суадносіны сапраўдных спрацоўванняў (TP) да агульнай колькасці станоўчых прагнозаў (TP + ілжывыя спрацоўванні (FP)). Вымярае долю станоўчых прагнозаў, якія на самой справе дакладныя. Высокая дакладнасць азначае, што мадэль робіць менш ілжывых дадатковых прагнозаў;

водгук (recall) – гэта суадносіны сапраўды станоўчых вынікаў (TP) да агульнай колькасці фактычна станоўчых выпадкаў (TP + ілжываадмоўныя вынікі (FN)). Адзначае долю рэальных дадanych выпадкаў, правільна ідэнтыфікаваных мадэллю. Высокая паўната азначае, што мадэль робіць менш ілжываадмоўных прагнозаў;

верагоднасць (accuracy) – гэта стаўленне колькасці правільных прагнозаў (TP + TN) да агульнай колькасці прагнозаў (TP + TN + FP + FN). Ацэньвае агульную прадукцыйнасць мадэлі, улічваючы як правільныя, так і няправільныя прагнозы;

забеспячэнне (support) паказвае, колькі ўзораў (адрэзкаў спектраграм) трапіла ў тэставую выбарку па той ці іншай птушцы.

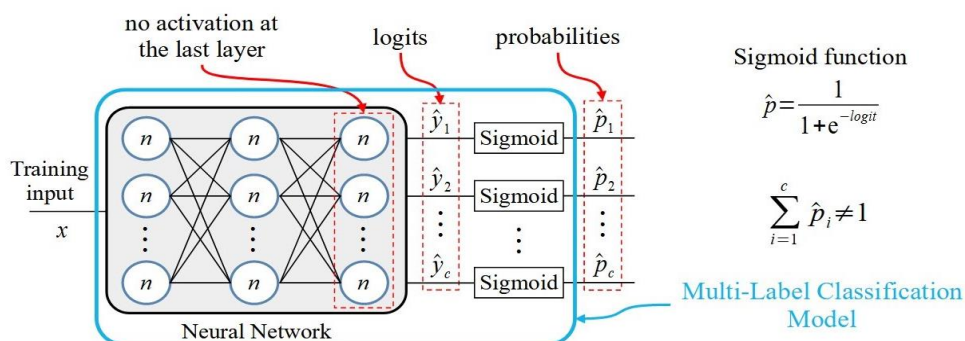
Такім чынам, дакладнасць вымярае якасць дадатных прагнозаў мадэлі, водгук – ахоп дадатнымі прагнозамі мадэлі, а верагоднасць – агульную правільнасць прагнозаў мадэлі. f1-score – сярэдняе гарманічнае значэнне паміж дакладнасцю і водгукам. Гэтыя метрыкі могуць прымаць значэнні ў прамежку ад 0 да 1, дзе значэнне 0 паказвае, што мадэль заўсёды памыляецца, а 1, – што мадэль заўсёды прадказвае правільна. Такім чынам, мадэль дэманструе наступныя паказчыкі:

accuracy: 0.5879;

macro avg: f1-score 0.5216, recall 0.5311, precision 0.5758;

weighted avg: f1-score 0.5473, recall 0.5879, precision 0.5952.

Пасля тэсціравання мадэлі Multi-Class Classification было прынята рашэнне прымянення іншай мадэлі навучання, менавіта *Multi-Label Classification*, паколькі папярэдняя мадэль распазнавання не ажыццяўляе аналіз птушыных хораў. Для атрымання імавернасці прыналежнасці гука ў акне да пэўнага класа і прадказання наяўнасці некалькіх відаў адначасова абрана функцыя актывацыі *Sigmoid*. Задача Multi-Label Classification ажыццяўляе класіфікацыю ўваходных даных на класы колькасцю больш за 2 незалежна адзін ад аднаго. Функцыя актывацыі *Sigmoid* падыходзіць да гэтай задачы, таму што размяркоўвае верагоднасці, каб кожная з іх была ў прамежку ад 0 да 1 па кожнаму класу незалежна. Актыватар *Sigmoid* ставіць у адпаведнасць аднаму акну адначасова некалькі класаў ці нічога. Схема атрыманай архітэктury мадэлі Multi-Label Classification прыведзена на мал. 6, а вынікі яе працы – на мал. 7 [18].

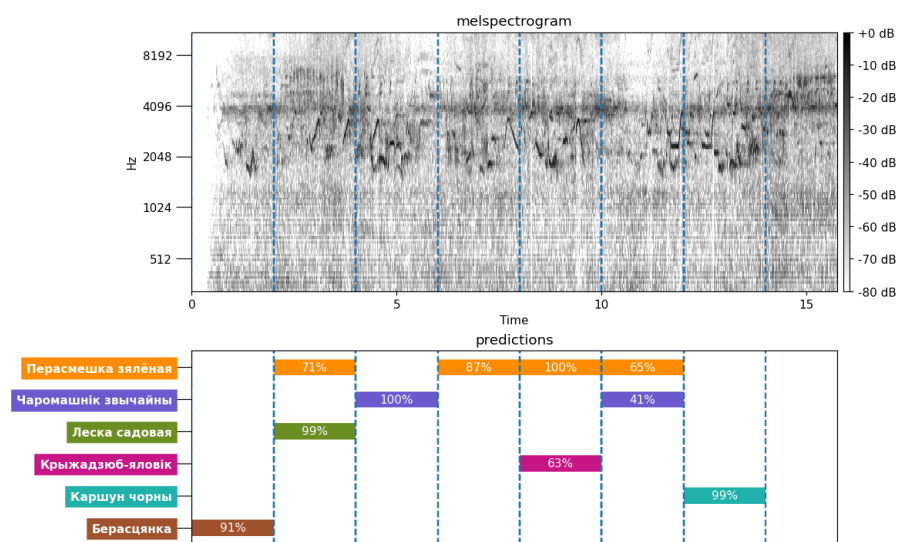


Мал. 6. Архітэтура мадэлі Multi-Label Classification з функцыяй актывацыі Sigmoid

Fig. 6. Architecture of the Multi-Label Classification model with Sigmoid activation function

Для ацэнкі навучання мадэлі Multi-Label Classification выкарыстоўваюцца метрыкі дакладнасць, водгук і f1-score. Метрыка верагоднасць у такіх задачах, дзе адно акно можа належыць некалькім класам, з'яўляецца непаказальнай. Гэта звязана з тым, што яна падлічвае супадзенні па ўсім класам, нават па тым, якіх няма на запісе, аднак ёсць у спісе відаў. Таму метрыкі дакладнасць, водгук і f1-score даюць больш падрабязную інфармацыю па навучанні мадэлі. Вынікі навучання і тэсціравання мадэлі на ўсім масіве даных галасоў птушак прадэманстраваны ніжэй:

	precision	recall	f1-score	support
micro avg	0.53	0.52	0.52	10367
macro avg	0.52	0.51	0.49	10367
weighted avg	0.55	0.52	0.51	10367
samples avg	0.53	0.56	0.52	10367



Мал. 7. Спектраграма анатавання мадэлі Multi-Label Classification з адлюстраваннем верагоднасцей

Fig. 7. An annotation spectrogram of the Multi-Label Classification model with probability display

На вынікі навучання ўплываюць колькасць даных, колькасць відаў для прадказання, складанасць прадказання некаторых птушак з-за іх падобнасці ці здольнасці імітаваць розныя гукі, якасць запісаў і інш. Акрамя гэтага для мадэлі Multi-Label Classification патрэбна больш даных, чым для мадэлі Multi-Class Classification з-за незалежнага прадказання кожнага з відаў [19]. Улічваючы гэтыя моманты, атрыманая мадэль паказвае даволі добрыя вынікі прадка-

зання на бягучым этапе. У далейшым плануецца павялічваць колькасць даных для навучання, каб палепшыць навучанне мадэлі.

Макет праграмнага забеспячэння аўтаматызаванай разметкі галасавых сігналаў птушак (жывёл) у сабраных базах даных. У выніку праведзеных прац распрацаваны і пратэсціраваны функцыянал забеспячэння аўтаматызаванай разметкі галасавых сігналаў птушак (жывёл) у сабраных базах даных для эксперыментальнага праграмнага забеспячэння аўтаматызаванага распазнавання галасавых сігналаў птушак (жывёл). Ён цалкам інтэграваны ў інфармацыйна-аналітычны цэнтр бесперапыннага маніторынгу, які знаходзіцца ў адкрытым доступе². Інфармацыйна-аналітычны цэнтр бесперапыннага маніторынгу ўяўляе сабой сістэму, якая складаецца з такіх кампанентаў, як праграмнае забеспячэнне баз даных, масіў даных электронных карпусоў для навучання мадэлей распазнавання, вэб-сайт, які дапамагае загрузаць і апрацоўваць даныя з іншых крыніц і генеруе вынікі распазнавання [20].

Вынікам працы праграмнага забеспячэння з’яўляюцца пабудаваная мел-спектраграма аўдыязапісу і аўтаматызаваная разметка галасавых сігналаў жывёл на падставе апрацаванага аўдыязапісу. Аўтаматызаваную разметку магчыма згенераваць у тэкставы файл. Вартасць распрацаванага забеспячэння заключаецца ў простым узаемадзеянні карыстальніка з інтэрфейсам, вялікай навуковай базы даных галасоў птушак з бесперапынным абнаўленнем, а таксама выкарыстаннем прыкладання як рэсурса для атрымання новых ведаў. Інтэрфейс прататыпа інтуітыўна зразумелы, зручны і прасты ў выкарыстанні. Наведвальнікі сайта знойдуць патрэбную інфармацыю або выканаюць свае асабістыя задачы.

Прыведзены набор тэхналогій для фронт-энду і бэк-энду распрацоўкі праграмнага забеспячэння, фрэймворкі, базы даных і мовы праграмавання прыменены для рэалізацыі прыкладання. Стварэнне сервернай часткі вэб-прыкладання адбываецца з дапамогай наступных кампанентаў:

MySQL – база даных, якая выкарыстоўваецца для захавання інфармацыі пра карыстальнікаў, аўдыя, даных распазнавання і іншых даных на сайце;

REST API – дазваляе ўзаемадзейнічаць з платформай праз знешнія прыкладанні і сэрвісы;

Django – Python – фрэймворк для распрацоўкі сервернай часткі сайта;

Git – размеркаваная сістэма кантролю версій, якая выкарыстоўваецца для кіравання зменамі, сумеснай працы і адсочвання кода;

Nginx/Apache – вэб-серверы, якія апрацоўваюць і аддаюць вэб-старонкі;

Docker – тэхналогія кантэйнерызацыі для арганізацыі ізаляванай прасторы на серверы;

DNS – інфраструктура, якая задзейнічана для прывязкі дамена да вэб-серверу і забеспячэння даступнасці сайта ў Інтэрнэце;

SSL/TLS – пратаколы шыфравання, якія забяспечваюць бяспечную перадачу даных паміж кліентамі і серверам.

Распрацоўка і выкарыстанне мадэлі матэматычнага алгарытму распазнавання галасоў птушак ажыццяўлялася праз бібліятэкі:

Tensorflow (для машыннага навучання, з дапамогай якой ствараюцца і трэніруюцца нейронныя сеткі).

Librosa (для апрацоўкі і аналізу аўдыяданых для стварэння мел-спектраграм).

Pandas (бібліятэка ў Python для аналізу і апрацоўкі структураваных даных, такіх як звесткі пра вокны для навучання).

NumPy (бібліятэка ў Python для матэматычных вылічэнняў з шматмернымі масівамі, з яе дапамогай захоўваюцца мел-спектраграмы ў фармаце *.npy*).

Matplotlib (для візуалізацыі даных для мел-спектраграм і будавання графікаў для статьистыкі).

Scikit-learn (для раздзялення даных на выбаркі для трэніроўкі і падліка вынікаў навучання).

Keras (адкрытая бібліятэка на Python, прыменена сумесна з *Tensorflow* для працы з мадэлямі нейронных сетак).

²Інфармацыйна-аналітычны цэнтр бесперапыннага аўтаматызаванага маніторынгу рэдкіх, пагражальных і індыкатарных відаў жывёл (птушак) : [сайт]. – 2024. – URL: <https://bird-voice-iac.ssrilab.by/> (дата звароту: 06.06.2024).

Спіс бібліятэк, якія прыменены для рэалізацыі фронт-часткі вэб-сайта:

1) HTML/CSS – мовы разметкі і стыляў выкарыстоўваюцца для стварэння вонкавага выгляду і макета вэб-старонак;

2) JavaScript – кліенцкая тэхналогія праграмавання для стварэння інтэрактыўных элементаў, анімацыі і іншых сцэнарыяў на вэб-сайце;

3) formik – для кіравання формамі ў react, дазваляе лёгка ствараць, валідаваць і адпраўляць формы;

4) i18next – для інтэрнацыяналізацыі (i18n), адказвае за рэалізацыю шматмоўнай падтрымкі для карыстальніцкага інтэрфейсу;

5) react – асноўная бібліятэка для стварэння карыстальніцкіх інтэрфейсаў.

Таксама прымяняюцца такія бібліятэкі, як react-dom, react-i18next, react-redux, react-router-dom, react-scripts, sass, swiper, web-vitals, yup, wavesurfer.js. Прыведзеныя рэсурсы неабходны для прадстаўлення сістэмы карыстальніку, забеспячэння якаснай рэалізацыі функцыянала, які адпавядае сучасным тэхналогіям, зручнай працы для дасягнення мэты праекта.

Заклучэнне. Даследаванне галасавых сігналаў птушак і распрацоўка сістэм аўтаматычнага распазнавання птушыных галасоў аказваюць значны ўплыў на маніторынг біразнастайнасці, ахову фаўны і экасістэм. Выкарыстанне праграмага забеспячэння, заснаванага на матэматычных мадэлях, якія ўключаюць нейронныя сеткі і іх глыбокае машыннае навучанне, дазваляе ствараць больш дакладныя і эфектыўныя мадэлі аналізу галасавых сігналаў птушак і іх генерацыі.

Падчас правядзення даследавання па распрацоўцы тэхналогіі аўтаматызаванага распазнавання галасавых сігналаў птушак (жывёл) для аўтаномнага бесперапыннага маніторынгу рэдкіх, пагражальных і індыкатарных відаў, а таксама стану біразнастайнасці ў лясных экасістэмах былі дасягнуты значныя поспехі і адкрыты перспектывы для далейшага развіцця ў гэтай галіне. У ходзе распрацоўкі метадычных асноў для збору, анатавання і генерацыі галасавых сігналаў жывёл на тэрыторыі Беларусі была створана база даных, скампіляваная з розных крыніц. Яна складаецца з больш за 2500 аўдыязапісаў для 116 відаў птушак, з якіх атрымана каля 42 000 анатаваных вакалізацый. Прыведзены набор даных уключае запісы рознай якасці, а таксама аўдыя з рознай колькасцю сумесна вакалізуючых відаў. Методыка анатавання аўдыяфайлаў для аўтаматызаванага распазнавання ўключае іх разбіццё на вокны неабходнага памеру, камп'ютарны аналіз і пабудову мел-спектраграмы для кожнага файла.

Для распрацоўкі эксперыментальнага праграмага забеспячэння аўтаматычнага распазнавання галасавых сігналаў птушак (жывёл) рэалізаваны работы па стварэнні мадэлі і алгарытмаў яго функцыянавання. У аснове мадэлі ляжыць перадавацкая скруткавая нейронная сетка з дадатковымі пластамі Flatten, Dropout і Dense. У якасці функцыі актывацыі для апошняга пласта і функцыі страт былі абраны Softmax і Sigmoid, для функцыі страт – бінарная крыжаваная энтрапія (binary cross entropy), і задача з вобласці Multi-Class Classification (адно акно – адно прадказанне) перайшла ў Multi-Label Classification (адно акно – некалькі прадказанняў). Працэс навучання нейрасеткі рэалізаваны на 50 эпохах са змяненнем хуткасці навучання (learning rate) па ReduceLROnPlateau.

Ацэнка працы мадэлі ажыццяўлялася праз выкарыстанне метрык дакладнасць, водгук і іх сярэдняга геаметрычнага значэння f1-score. Для тэсціравання мадэлі са скампіляванага масіва даных складзена выбарка такім чынам, каб адзін і той жа запіс не трапіў адначасова ў розныя выбаркі. У выніку праверкі на абраных 10 000 анатацый тэставых даных мадэль Multi-Class Classification паказала сярэднія значэнні precision 0.59, recall 0.58 і f1-score 0.54 адпаведна, а мадэль Multi-Label Classification прадэманстравала сярэднія значэнні precision 0.53, recall 0.56 і f1-score 0.52 адпаведна, што лічыцца прымальным вынікам на дадзеным этапе распрацоўкі мадэлі.

У выніку праведзеных прац быў распрацаваны і паспяхова пратэсціраваны макет праграмага забеспячэння для аўтаматычнага распазнавання галасавых сігналаў птушак. Праграмага забеспячэнне цалкам інтэгравана ў інфармацыйна-аналітычны цэнтр бесперапыннага маніторынгу і даступна для выкарыстання ў адкрытым доступе. Такім чынам, выкананыя задачы не толькі далі значныя навуковыя і тэхнічныя вынікі, але і адкрылі шляхі для далейшага ўдасканалення метадаў маніторынгу і абароны фаўны Беларусі, спрыяючы захаванню і вывучэнню

біразнастайнасці ў лясных экасістэмах. Атрыманыя вынікі дазваляць распаўсюдзіць веды і вопыт у арніталогіі сярод шырокай аўдыторыі і правесці далейшыя даследаванні ў гэтым кірунку.

Уклад аўтараў. Ю. С. Гецэвіч прапанаваў канцэпцыю аўтаматызаванага распазнавання галасавых сігналаў птушак, ажыццяўляў абмеркаванне і кантроль выканання этапаў распрацоўкі праграмага забеспячэння, тэсціраваў сістэму на мэтавую адпаведнасць. А. А. Бакуновіч распрацаваў пакрокавы алгарытм працы вэб-прыкладання і ажыццявіў яго рэалізацыю на API, унёс карэктарыі пасля тэсціравання праграмага забеспячэння. Т. Г. Шагава і Д. А. Жалава займаліся распрацоўкай канцэпцыі алгарытма матэматычнай мадэлі распазнавання, праводзілі эксперыменты праверкі гіпотэз паляпшэння алгарытма, пабудавалі сістэмы бесперапыннага доступу да мадэляў, наладжвалі асяроддзі для навучання і непасрэднага запуску працэсаў навучання, збіралі масіў даных і ажыццяўлялі іх перадапрацоўку. Я. С. Зяноўка вяла працы па афармленню навуковых апісанняў праекта, ажыццяўляла аналіз тэарэтычных і практычных вынікаў пакрокавай рэалізацыі праекта.

Спіс выкарыстаных крыніц

1. Priyadarshani, N. Automated birdsong recognition in complex acoustic environments: a review / N. Priyadarshani, S. Marsland, I. Castro // *Journal of Avian Biology*. – 2018. – URL: <https://nsojournals.onlinelibrary.wiley.com/doi/full/10.1111/jav.01447> (date of access: 19.03.2021).
2. Sharma, S. A methodological literature review of acoustic wildlife monitoring using artificial intelligence tools and techniques / S. Sharma, K. Sato, B. P. Gautam // *Sustainability*. – 2023. – Vol. 15, no. 9. – P. 7128. – DOI: 10.3390/su15097128.
3. Briggs, F. Audio classification of bird species: a statistical manifold approach / F. Briggs, R. Raich, X. Z. Fern // 2009 Ninth IEEE Intern. Conf. on Data Mining, Miami Beach, FL, USA, 6–9 Dec. 2009. – Miami Beach, 2009. – P. 51–60.
4. BirdNET: A deep learning solution for avian diversity monitoring / S. Kahl, C. M. Wood, M. Eibl, H. Klinck // *Ecological Informatics*. – March 2021. – Vol. 61. – P. 101236. – DOI: 10.1016/j.ecoinf.2021.101236.
5. Insights and approaches using deep learning to classify wildlife / Z. Miao, K. M. Gaynor, J. Wang [et al.] // *Scientific Reports*. – 2019. – Vol. 9, no. 1. – URL: <https://www.nature.com/articles/s41598-019-44565-w> (date of access: 13.02.2021).
6. Тэхналогіі аўтаматычнай апрацоўкі і аналізу маўлення з прымяненнем штучнага інтэлекту / Ю. С. Гецэвіч, В. В. Дыдо, Д. А. Бяляўскі [і інш.] // II Форум IT-Академграда «Искусственный интеллект в Беларуси»: доклады, Минск, 12–13 окт. 2023 г. – Минск: ОИПИ НАН Беларуси, 2023. – С. 71–78.
7. Klein, D. J. Deep learning for large scale biodiversity monitoring / D. J. Klein, M. Mckown, B. Tershy // *Bloomberg Data for Good Exchange Conf.*, N. Y., NY, USA, 28 Sept. 2015. – N. Y., 2015. – 7 p. – DOI: 10.13140/RG.2.1.1051.7201.
8. Artificial intelligence (BirdNET) supplements manual methods to maximize bird species richness from acoustic data sets generated from regional monitoring / L. Ware, C. L. Mahon, L. McLeod, J. F. Jetté // *The Canadian Journal of Zoology*. – 2023. – Vol. 101, no. 12. – P. 1031–1051.
9. Stowell, D. An open dataset for research on audio field recording archives: freefield1010 / D. Stowell, M. D. Plumbley. – 2013. – URL: <https://arxiv.org/abs/1309.5275> (date of access: 06.06.2024).
10. The machine learning-powered BirdNET App reduces barriers to global bird research by enabling citizen science participation / C. M. Wood, S. Kahl, A. Rahaman, H. Klinck // *PLoS Biology*. – 2022. – Vol. 20, no. 6. – 10 p. – DOI: 10.48550/arXiv.1309.5275.
11. The Belarusian list of bird species approved by the Belarusian ornitho-faunistic commission for 2021 and 2022 / N. V. Karlionova, A. V. Borodin, I. E. Samusenko, M. Y. Nikiforov // *Zoological Readings*. – Grodno: GrGU, 2023. – P. 113.
12. Мадэль баз даных для тэхналогіі аўтаматызаванага распазнавання галасавых сігналаў жывёл / С. А. Гайдурэў, Д. І. Латышэвіч, А. А. Бакуновіч [і інш.] // Развитие информатизации и государственной системы научно-технической информации (РИНТИ-2022): докл. XXI Междунар. науч.-техн. конф., Минск, 17 нояб. 2022 г. – Минск: ОИПИ НАН Беларуси, 2022. – С. 236–240.
13. Stowell, D. Computational bioacoustics with deep learning: a review and roadmap / D. Stowel // *PeerJ*. – 2022. – Vol. 10 – P. 46.

14. FSD50K: an open dataset of human-labeled sound events / E. Fonseca, X. Favory, J. Pons [et al.] // *ACM Transactions on Audio, Speech, and Language Processing*. – 2022. – Vol. 30 – P. 829–852.
15. Hearing to the unseen: audiomoth and BirdNET as a cheap and easy method for monitoring cryptic bird species / G. Bota, R. Manzano-Rubio, L. Catalán [et al.] // *Sensors*. – 2023. – Vol. 23, no. 16. – 11 p. – DOI: 10.3390/s23167176.
16. Tan, M. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks / M. Tan, Q. V. Le // *Proc. of the 36th Intern. Conf. on Machine Learning, ICML 2019, Long Beach, 9–15 June 2019*. – Long Beach, 2019. – P. 6105–6114.
17. Multi-class imbalanced data classification: a systematic mapping study / Y. Wang, M. M. Rosli, N. Musa, F. Li // *Engineering, Technology & Applied Science Research*. – 2024. – Vol. 14. – P. 14183–14190. – DOI: 10.48084/etasr.7206.
18. Mao, J.-X. Learning label-specific multiple local metrics for multi-label classification / J.-X. Mao, J.-Y. Hang, M.-L. Zhang // *Thirty-Third Intern. Joint Conf. on Artificial Intelligence {IJCAI-24}*, Jeju, Korea, 3–9 Aug. 2024. – Jeju, 2024. – P. 4742–4750. – DOI: 10.24963/ijcai.2024/524.
19. Jia, B.-B. Towards exploiting linear regression for multi-class/multi-label classification: an empirical analysis / B.-B. Jia, J.-Y. Liu, M.-L. Zhang // *International Journal of Machine Learning and Cybernetics*. – March 2024. – Vol. 15. – P. 3671–3700. – DOI: 10.1007/s13042-024-02114-6.
20. Developing birds sound recognition system using an ontological approach / Ya. Zianouka, D. Bialiauski, L. Kajharodava [et al.] // *Open Semantic Technologies for Intelligent Systems*. – Minsk, Belarusian State University of Informatics and Radioelectronics, 2023. – Iss. 7. – P. 165–170.

References

1. Priyadarshani N., Marsland S., Castro I. Automated birdsong recognition in complex acoustic environments: a review. *Journal of Avian Biology*, 2018, vol. 49, no. 5. Available at: <https://nsojournals.onlinelibrary.wiley.com/doi/full/10.1111/jav.01447> (accessed 19.03.2021).
2. Sharma S., Sato K., Gautam B. P. A methodological literature review of acoustic wildlife monitoring using artificial intelligence tools and techniques. *Sustainability*, 2023, vol. 15, no. 9, p. 7128. DOI: 10.3390/su15097128.
3. Briggs F., Raich R., Fern X. Z. Audio classification of bird species: a statistical manifold approach. *2009 Ninth IEEE International Conference on Data Mining, Miami Beach, FL, USA, 6–9 December 2009*. Miami Beach, 2009, pp. 51–60.
4. Kahl S., Wood C. M., Eibl M., Klinck H. BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, vol. 61, March 2021, p. 101236. DOI: 10.1016/j.ecoinf.2021.101236.
5. 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, no. 1. Available at: <https://www.nature.com/articles/s41598-019-44565-w> (accessed 13.02.2021). DOI: 10.1038/s41598-019-44565-w.
6. Hetsevich Yu. S., Dydo O. V., Bialiauski D. A., Zjanowka Ja. S., Ljucich M. S., ..., Nazaraw U. U. *Technologies of automatic speech processing and analysis using artificial intelligence*. Doklady II Forumu IT-Akademgrada «Ikskusstvennyj intelekt v Belarusi» [Reports of the II IT Academy Forum "Artificial Intelligence in Belarus", Minsk, 12–13 October 2023]. Minsk, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus, 2023, pp. 71–78 (In Bel).
7. Klein D. J., Mckown M. W., Tershy B. R. Deep learning for large scale biodiversity monitoring. *Bloomberg Data for Good Exchange Conference, New York, NY, USA, 28 September 2015*. New York, 2015, 7 p. DOI: 10.13140/RG.2.1.1051.7201.
8. Ware L., Mahon C. L., McLeod L., Jetté J. F. Artificial intelligence (BirdNET) supplements manual methods to maximize bird species richness from acoustic data sets generated from regional monitoring. *The Canadian Journal of Zoology*, 2023, vol. 101, no. 12, pp. 1031–1051.
9. Stowell D., Plumbley M. D. *An open dataset for research on audio field recording archives: freefield1010*, 2013. Available at: <https://arxiv.org/abs/1309.5275> (accessed 06.06.2024). DOI: 10.48550/arXiv.1309.5275.
10. Wood C. M., Kahl S., Rahaman A., Klinck H. The machine learning-powered BirdNET App reduces barriers to global bird research by enabling citizen science participation. *PLoS Biology*, 2022, Vol. 20, no. 6, 10 p. DOI: 10.48550/arXiv.1309.5275.
11. Karlionova N. V., Borodin A. V., Samusenko I. E., Nikiforov M. Y. The Belarusian list of bird species approved by the Belarusian Ornitho-Faunistic Commission for 2021 and 2022. *Zoological Readings*, Grodno, Grodnenskij gosudarstvennyj universitet imeni Janki Kupaly, 2023, p. 113.

12. Gaidurov S. A., Latyshevich D. I., Bakunovich A. A., Kaigorodova L. I., Khokhlov V. A., ..., Hetsevich Yu. S. *A database model for automated recognition of animal voice signals*. Razvitie informatizacii i gosudarstvennoj sistemy nauchno-tehnicheskoy informacii RINTI-2022 : doklady XXI Mezhdunarodnoj nauchno-tehnicheskoy konferencii, Minsk, 17 nojabrja 2022 g. [*Development of Informatization and the State System of Scientific and Technical Information (RINTI-2022) : Reports of the XXI International Scientific and Technical Conference, Minsk, 17 November 2022*], The United Institute of Informatics Problems of the National Academy of Sciences of Belarus, 2022, pp. 236–240 (In Bel.).

13. Stowell D. Computational bioacoustics with deep learning: a review and roadmap. *PeerJ*, 2022, vol. 10, p. 46.

14. Fonseca E., Favory X., Pons J., Font F., Serra X. FSD50K: an open dataset of human-labeled sound events. *ACM Transactions on Audio, Speech, and Language Processing*, 2022, vol. 30, pp. 829–852.

15. Bota G, Manzano-Rubio R., Catalán L., Gómez-Catasús J., Pérez-Granados C. Hearing to the unseen: audiomoth and BirdNET as a cheap and easy method for monitoring cryptic bird species. *Sensors*, vol. 23, no. 16, 11 p. DOI: 10.3390/s23167176.

16. Tan M., Le Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Long Beach, 9–15 June 2019*. Long Beach, 2019, pp. 6105–6114.

17. Wang Y., Rosli M. M., Musa N., Li F. Multi-class imbalanced data classification: a systematic mapping study. *Engineering, Technology & Applied Science Research*, 2024, vol. 14, pp. 14183–14190. DOI: 10.48084/etasr.7206

18. Mao J.-X., Hang J.-Y., Zhang M.-L. Learning label-specific multiple local metrics for multi-label classification. *Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24), Jeju, Korea, 3–9 Aug. 2024*. Jeju, 2024, pp. 4742–4750. DOI: 10.24963/ijcai.2024/524.

19. Jia B.-B., Liu J.-Y., Zhang M.-L. Towards exploiting linear regression for multi-class/multi-label classification: an empirical analysis. *International Journal of Machine Learning and Cybernetics*, March 2024, vol. 15, pp. 3671–3700. DOI: 10.1007/s13042-024-02114-6.

20. ZianoukaYa., Bialiauski D., Kajharodava L., Chachlou V., Hetsevich Yu., ..., Zhakhsylyk 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, iss. 7, pp. 165–170.

Інфармацыя пра аўтараў

Геўзэвіч Юрась Станіслававіч, кандыдат тэхнічных навук, дацэнт, загадчык лабараторыі распазнавання і сінтэзу маўлення, Аб'яднаны інстытут праблем інфарматыкі Нацыянальнай акадэміі навук Беларусі.
E-mail: yuras.hetsevich@gmail.com

Зяноўка Яўгенія Сяргеёўна, малодшы навуковы супрацоўнік, Аб'яднаны інстытут праблем інфарматыкі Нацыянальнай акадэміі навук Беларусі.
E-mail: evgeniakacan@gmail.com

Бакуновіч Андрэй Аляксеевіч, малодшы навуковы супрацоўнік, Аб'яднаны інстытут праблем інфарматыкі Нацыянальнай акадэміі навук Беларусі.
E-mail: bakunovich.andrei.work@gmail.com

Жалава Дар'я Аляксандраўна, інжынер-праграміст, Аб'яднаны інстытут праблем інфарматыкі Нацыянальнай акадэміі навук Беларусі.
E-mail: daryazhalova@gmail.com

Шагава Таццяна Грыгор'еўна, інжынер-праграміст, Аб'яднаны інстытут праблем інфарматыкі Нацыянальнай акадэміі навук Беларусі.
E-mail: tanya.shagova@gmail.com

Information about the authors

Yuras S. Hetsevich, Ph. D. (Eng.), Assoc. Prof., Head of the Speech Synthesis and Recognition Laboratory, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: yuras.hetsevich@gmail.com

Yauheniya S. Zianouka, Junior Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: evgeniakacan@gmail.com

Andrei A. Bakunovich, Junior Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: bakunovich.andrei.work@gmail.com

Zhalava Darja, Software Engineer, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: daryazhalova@gmail.com

Tatsiana G. Shagava, Software Engineer, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: tanya.shagova@gmail.com