CLASSIFY BIRD SPECIES AUDIO BY AUGMENT CONVOLUTIONAL NEURAL NETWORK

Hasan Abdullah Jasim
Electrical and computer engineering
Altinbas University
University of Tikrit
Istanbul, turkey
emperor.hasan@yahoo.com

Saadaldeen R. Ahmed
Computer science
University of Tikrit
Baghdad, Iraq Saa
Saadaldeen.r.ahmed @tu.edu.iq

Abdullahi Abdu IBRAHIM
Electrical and computer engineering
Altinbas University
Istanbul, turkey
abdullahi.ibrahim@altinbas.edu.tr

Adil Deniz DURU
Physical Education and Sports
Marmara University
Istanbul, turkey
deniz.duru@marmara.edu.tr

2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) | 978-1-6654-6835-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/HORA55278.2022.9799968

***Abstract*—Using convolutional neural networks, this thesis aims to create a system for fully automated identification of bird species based on spectrogram images. Spectrogram analysis is more difficult when trying to make an advance identification of a bird species. On a publicly available dataset of 8000 audio examples, we've begun by analyzing the challenges of bird species detection, segmentation, and classification to achieve our goal. It has been determined also that deep learning-based technique CNN with Fully convolutional learning calls for easier results because it eliminates the possible future modelling error caused by an imprecise knowledge of bird species and works well on coding in cohesion with the spectral analysis kernel using the librosa library. We have concluded. After obtaining the dataset from the open-source repository, it is then processed locally. For training, testing, and validation we used a subset of the dataset of 8000 sound samples. We offered a method relying on a CNN reset learned that proved to be very quick and optimum because it was first needing the spectrogram analytic kernel to learn what to class in bird species, and then it gets the system trained on features extracted. In a novel 9-step implementation, a bird species spectrogram can be detected from an audio sample. There was a loss of less than 0.0063, and the conditioning workouts accuracy is 0.9895 for the system, 0.9 as precision, and training and validation use 50 epochs in system.**

Keywords— Birds, xeno-canto, classification, CNN, audio, spectrogram, identification, librosa.

I. INTRODUCTION

As part of this dissertation, we used deep learning to detect and classify spectrogram stages of different bird species. For many years, the inability to accurately identify different species of birds has been a disturbing factor for bird watchers. Bird species of various kinds [1] are a common type found all over the world. Preventing bird species detection is possible, even though it's the most commonly fatal type of the problem [2]. Calculated tests have demonstrated that the division profit within every year make a recording, generally tracking people in an extraordinary-risk class, as stated in [2]. has been proven. Detection of birds' vocalizations can be done at an early stage during audio recording, thereby reducing the number of birds that are killed by the detection. To improve waterbirds detection and classification quality and efficiency, computer-assisted algorithms [3] are credited. It carries out a species evaluation and provides formalized reports on the volume, location, and other recommendations for more diagnosis and treatment. ' An automatic analysis system is provided by using CNN techniques to derive information from each scan. As mentioned in [4], it is a system that examines data from the method and uses its expertise to reach a final inference about potential detection via spectrogram.

Even with today's most advanced audio systems, there is always room for error. As a result of their sensitivity to non-bird species buildings like detection methods, many false-positive predictions are generated. Before presenting any candidates to a system, as described in [6], one step in the bird variety finding goal is to classify all regions that were garnered as animal data during image retrieval to get rid of unauthorized access. The ratio of false negatives to true positives varies enormously between the algorithms, and it is always the situation that many roles are discovered by the algorithms, but only a small percentage of those regions contain a spectrogram of a bird species. Separate classification models that learn to discern between spectra and non-spectrogram are used in [7] to improve the accuracy of a complete end-to-end detection system. There are many noises that can be examined by a system if they gotten probability for each noise sample is used as a filter.

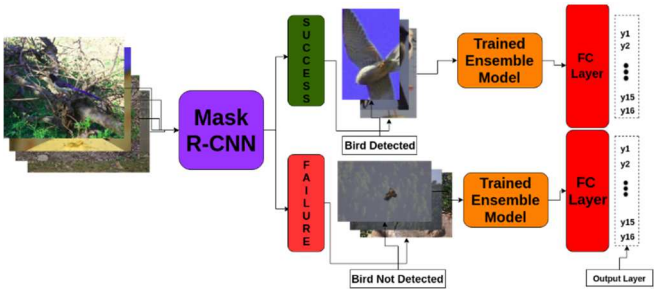

Fig. 1. Bird species division utilizing the transport learning [8].

Even today's most sophisticated CNN systems are unable to produce results free of errors. As a result, many erroneous positive predictions are produced when they are trained on architectures that do not include the bird species. Classifying all regions that were obtained as presidential hopefuls during ROI extraction is one step in bird species detection to reduce the amount of non-bird species before displaying all

candidates to a framework. Many detection techniques find many roles, but only a few contain a bird species, as described in [10]; this is the case with all detection algorithms. Models that learn to discern between bird and non-bird species are used to enhance the precision of a full end to end detection system. Once the regions that do not contain any bird species have been filtered out using the received likelihood for each ROI, a smaller set of images can be examined by a system.

The classification step has been approached in a variety of ways. As mentioned in [11], some approaches use classical machine learning approaches to extract features, followed by any type of classifier. The other methods rely on end-to-end DNN's that use convolution to complete the feature extraction process. The complexity of the data in imaging problems sets it apart from other areas of computer vision. But it is common for tasks that involve images of using methods and neural that were originally designed for other domains, such as image classification. Using 3-D deep CNN to solve the bird species classification problem is possible because the tasks data sources (scans) are three dimensional. It is also possible to lose information about a candidate's entire body by not using the entire volumetric region of data but instead representing it in a CNN [13].

Many new architectures for deep learning models are emerging. So, it is critical to assess which of these gives the best performance of the number of false positives produced and the time consumed [14]. [15, 16] The use of less complex methods could really preserve a similar outcome while saving some resources when DNN are an unnecessary overkill for a given task. Thus, thorough, and comprehensive studies are needed for each specific problem, which would compare various approaches to see which one is most effective as well as provide a report.

A. Research Contribution

• Highlight prior research and accomplishments in bird species detection, classification, or identification.

• Classify the regions using a unified dataset of bird species audio samples and three-dimensional deep learning techniques with an augmented CNN resent learning model.

• Furthermore, it is important to evaluate and compare the various convolutional neural networks that are currently in use.

• A new implementation pipeline to detect, analyze spectrograms and classify bird species should be developed in this regard.

II. LITERATURE REVIEW

Previous studies have used a variety of image features to pinpoint specific bird species' geographic ranges. In sound resonant frequency speeches (audio), the most common image feature is describing texture patterns as described in [19]. Gradient histograms as well as features calculated using the Discrete Wavelet Transform, as described in [20], make up most textural features. Birds and other reptilian organ abnormalities are typically classified using textural measurements. Multiple layer perceptron's (MLP), (SVM), and Artificial neural networks (ANN) are all common machine learning techniques, as described in [21].

The author of article [22] used gradient magnitude histograms, GLCM texture, and CNN analysis in subgroup ROIs of an MR image to identify bird species with cirrhosis. K-means clustering, an unsupervised method of learning, was used to achieve a sensitivity of 72% and specificity of 60%.

Using GLCM, DWT, and gray scale gradient co-occurrences, the writer in article [23] shows that wavelet transform and recognition of species diversity lesions in sound scans are valuable methodologies that produce good results. Many machine learning classification methods, including neural networks, SVMs and k-Nearest Neighbor, are evaluated, all of which do well in the tests conducted.

ADCNN and grey - level histograms are used by the authors of article [24] to detect cancer regions. Based on the principle of evolutionary biology, an ADCNN algorithm is used to find the best possible features. The ADCNN network uses the optimal parameter vector and achieves an accuracy of classification of 96.2 percent for breast cancer pixels. Deep learning kernel CNN-classifiers were able to deal with the imbalanced data at hand, as demonstrated by their AUC scores. For the identification and tracking process as described in [25], an immediate offset was observed without the need for thresholding of the classifiers.

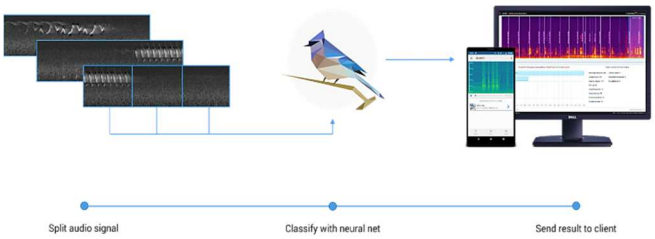

Fig. 2. The bird genus spectrograph using the NN by [26].

The survival of patients with non-bird audio identification and bird audio detection was examined separately in two studies by the authors in [27, 28]. ANNs were used in both studies to be educated using list of criteria using a forward stepwise selection method of feature selection and selection. The Models were found to be superior when compared to a variety of other scoring methodologies. With an utter AUC ROC value of 0.82, the best of these designs outperformed score-based approaches by up to 34%..

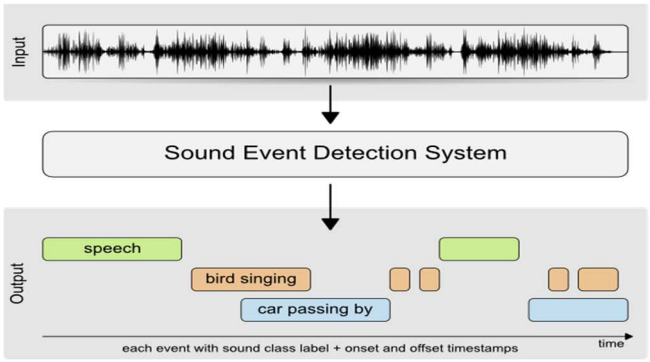

Fig. 3. Experience based noise recognition with class stamps and timestamps by Kaggle [29].

In many cutting-edge applications, such as [30], artificial neural networks (ANNs) produce high classification accuracy. The drawback is that from before the imaging ANNs are harder to come by and require more data to train than ANNs that are trained using natural images.

Based on previous research, the theory of machine learning-based measurements is tested with MLPs. In addition, because ANNs require longer training times, several

approaches can be tested within the allotted time for the thesis. Classification and regression of bird species using MLPs based on texture image features in two and three dimensions are discussed in [31]. . In research [32] We will look at various processes and approaches for early diagnosis of glaucoma using the MATLAB Deep Convolutional Neural Network (DCNN) [33] in this research. Many papers have begun to develop ways for boosting the efficiency of disease diagnosis, and one of the themes that has absorbed many publications is improving medical diagnostic methods. [34]. We claim 88.4 percent accuracy for the 5-row clustering challenge in this study [35]. For grouping the four groups aiming to discriminate cancer, we reported an accuracy of 92.3 percent and 96.2 percent, respectively, [36] and a sensitivity of 94.5 to 87.2 percent in the high-sensitivity action point [37]. According to everyone. The primary goal of this study is to determine the optimal tangential dark center point of tangent space using kinematic-based differential analysis of the brain-computer interface [38].

III. METHODOLOGY

The procedure was broken down into several steps. As part of this process, we created several modules, each with distinct responsibilities. Python programming was used to carry out the practical application of this dissertation research. Status to the sculpture appearing in the part, we take a closer look at some of the interesting theories that the CNN method has described for recognizing bird species identification on both static and dynamic levels:

1. Spectrogram analysis of the selected area is performed using the Data Acquisition module. Features like bird species localization, region saving and augmentation, and feature extraction are among them. We wrote these components with the intention of allowing their use at any point in the project. The data loading module, for example, can be used both during training and validation.

2. A module for each CNN-resent having to learn architects and spectrum analyzer analysis audio kernel has been written in Python language using the open-source machine learning libraries and answer options sound library (librosa) for modelling work. Since different environments have different configurations of audio libraries (resents) and kernels, it can teach models in each of them. This helps to compare the performance of the frameworks and to be more flexible when building the NN or changing existing ones.

3. A CNN-resent attempting to learn syntax on 10600 species of bird's audio samples was used to train and test two separate modules. It is possible to run a test validation step independent of model training to see the implementation pipeline as it is being built.

4. Pipeline: All the pre - processing that was done for the spectrogram classification was used in this step. It analyzes the extremely high of a system at various stages for species of bird detection and segmentation by taking the parameter estimates for each sound-based sample and running the analysis

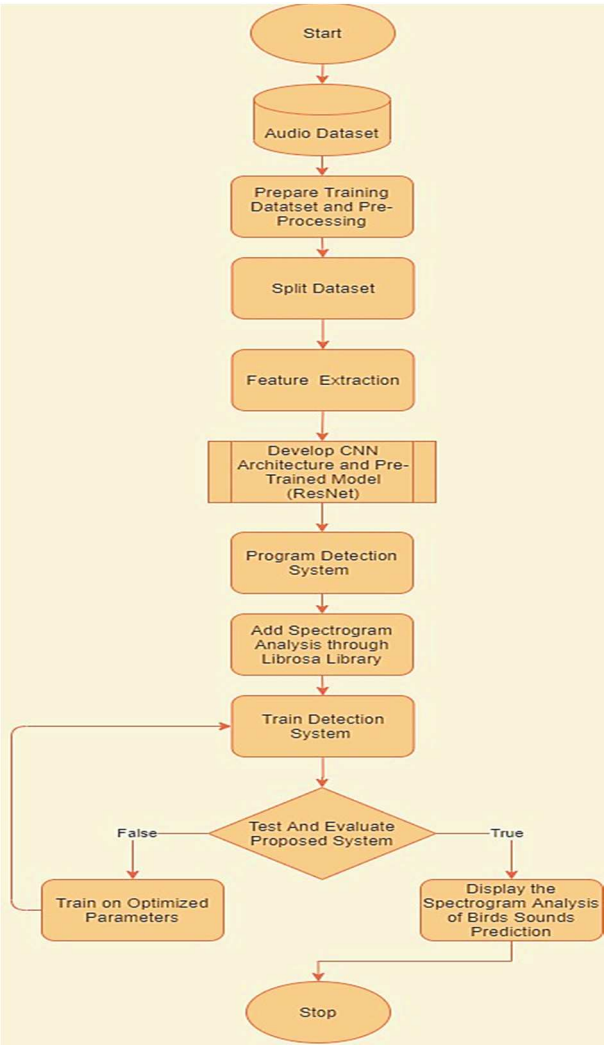

Fig. 4. Flow diagram explanation whole of the organizational part.

For bird species classification, we've developed a dependable app that makes use of an optimized pipeline of CNNs and bird's sounds, as well as CPU and GPU multicores. For the proposed application, the CPU's average processing time for a sound is 15.67 seconds. Modern GPU architecture uses an average of 3.57 seconds per sound. Bird species detection and classification is not affected by non-maxima suppression times from CNN transfer learning in this context again. Every minute is spent extracting sounds and birds, as well as gathering preliminary implementation pipelines, for use in developing the game. The CNN resent classification model is implemented using this application's execution pipeline. This phase of pipeline segmentation is relatively fast, but there are no proposed alternatives to bird's selective search regions. For example, using a set of 10600 set of data samples which have the same performance could yield a sound in seconds.

It sounds as if no one has ever done such a thing before, as the app was specifically created for the bird's species. Research in the field of convolutional and digital sound handling has benefited greatly from novel and practical applications. Bird species classification is a complex problem, and the most advanced tools for solving it are CNNs with sound processing that are extremely robust. Because of this application's three innovative python contributions, these CNNs with sound handling can handle images incredibly quickly.

IV. RESULTS

To make balancing CNN with resent teaching easier, three different types of training and verification data were recorded were extracted. There were examples of suitable spectrograms within bird species, as well as examples of unsuitable spectrograms within bird species. The performance of the extracted features from the NN is the most important factor in the existing system. CNN classifiers are focused on the complete coaching data set (8600 samples), tested just on test set (4500 sample were collected) to compare the power of different CNNs as well as feature layers. When data augmentation is used, it is possible that some corners of the lesion may be missing because of the random off-center cropping that is used. Real-time data increase over many epochs is a common method for training CNNs. As a result, the CNN will not only be able to assess the entire image, but it will also learn great functions from just a few patches of same lesion. The DCNN may have performed better with a more appropriate training input because of these results. It would be significant to continue these experiments, however, before drawing any definitive conclusions.

Using the validation data, we were able to accurately identify the bird species at the the last epoch of the classifier's training. The validation data's loss column displays the data's final average squared error function value. Several relative errors were considerably higher, it should be noted. Because of ROI false - positive in unsuitable positions and a small number of regions classified as suitable for birds, this occurred. Because of this, bird species tissue ROI's had smaller distributions, resulting in false bird's fraction values. So, some labels have been found to be more accurate enough, particularly when there are various skin areas in a single image that can be seen as cancerous and the comparison between lesion and surrounding skin isn't high enough. So, some labels are generated by a bird's intelligence rather than an algorithm, this scenario is possible. To put it another way, birds' inability to segment sounds means that certain segmentation labels do not accurately depict the shape of bird species. The results of detection and classification show that the extracted features outperform the segmentation tags in many cases. at these instances, index number is never incredibly high, but C.N.N achievement is extremely high because the division is within the neighborhood to the tags in Figure 5 The spectrogram testing and photographic description of field to incidence to bird varieties utilizing CNN type.

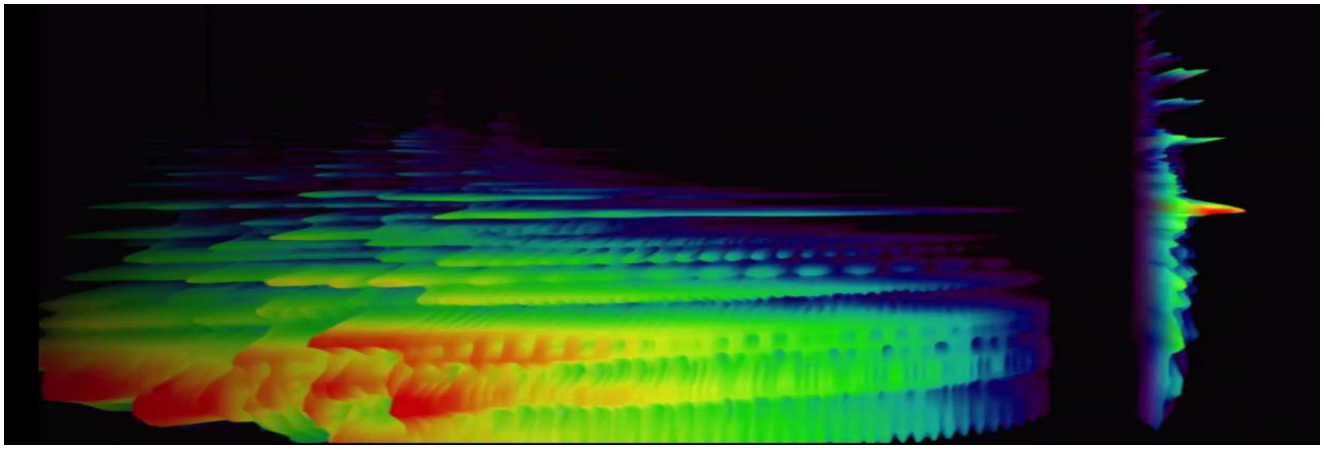

Fig 5. The spectrogram assessment and graphic description of field of incidence to bird varieties utilizing CNN type.

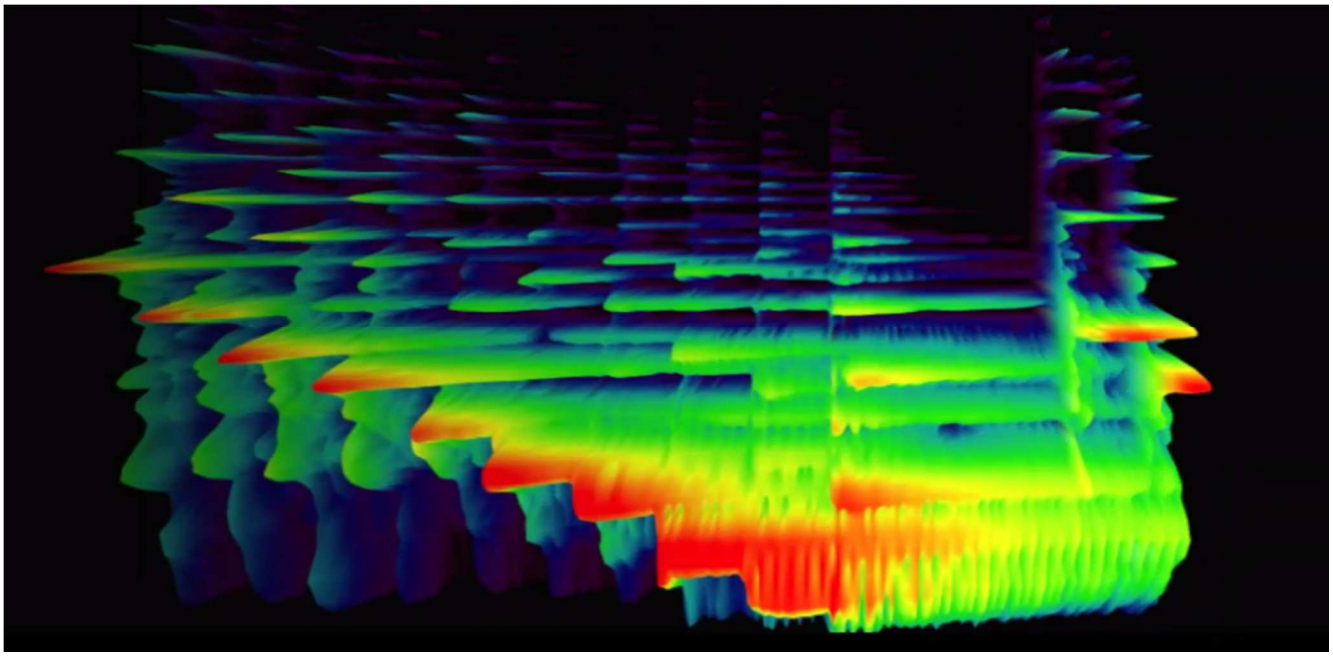

Fig 6. The spectrogram assessment and graphic description of field of incidence of bird varieties utilizing CNN type.

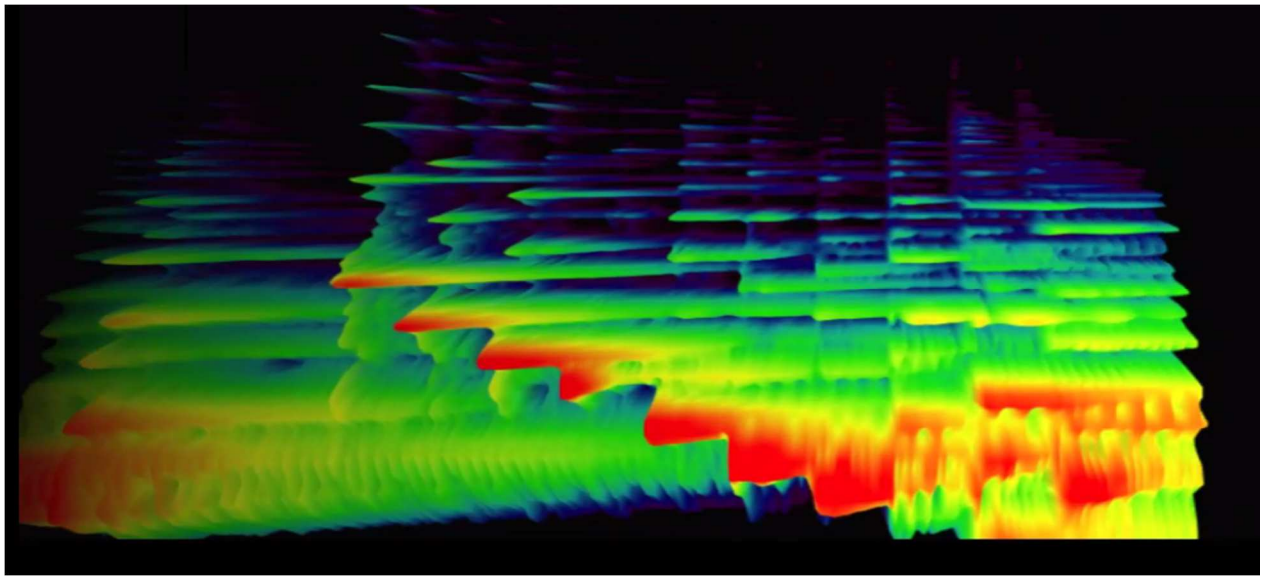

Fig 7. The spectrogram assessment and photographic description of field of incidence of bird varieties utilizing CNN type.

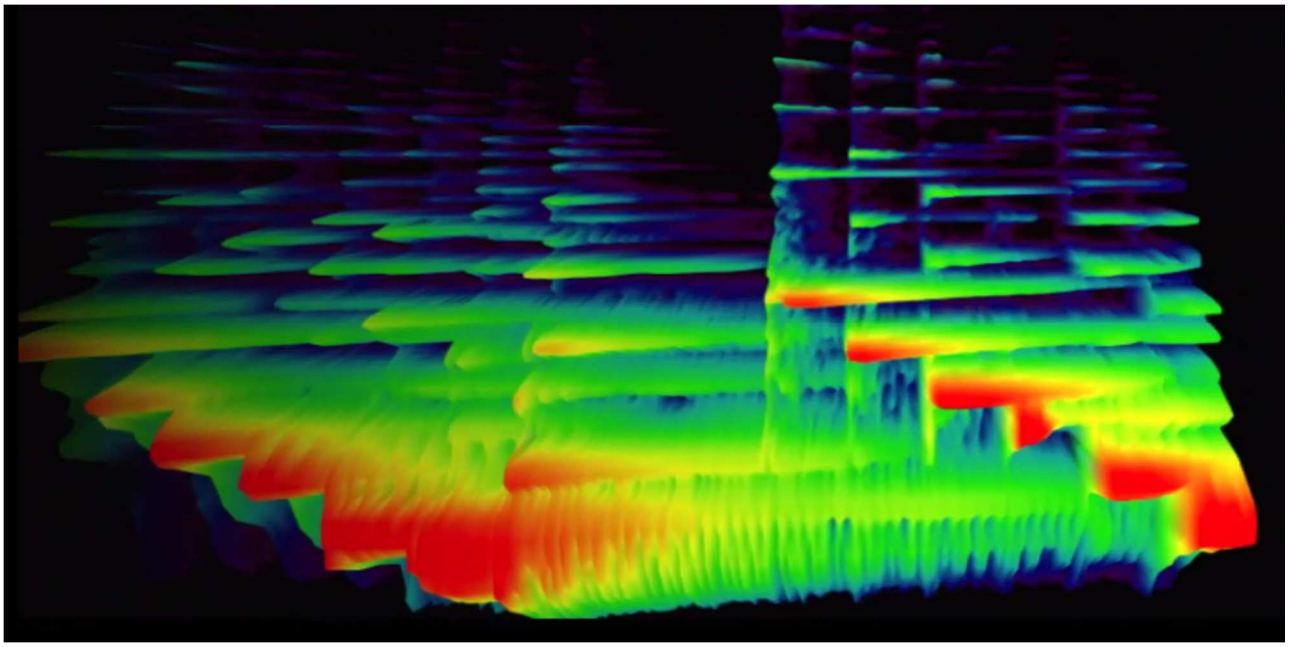

Fig 8. The spectrogram assessment and photographic depiction of continuum of incidence of bird varieties utilizing CNN type.

V. Conclusion

This study uses deep learning to build a high-accuracy model for bird audio. The experiment was solid, as which included a detailed analysis of bird aural sorter. The simulation stayed evaluated on top of (8000) audio sample. Content management was constantly monitored as the website's popularity grew and more multimedia content was uploaded. Many datasets have recently become public, inspiring this research. This study's framework uses CNN deep network properties to design an architecture that aids in features extraction. There will be a CNN with multiple convolution layers and a hybrid model. The hybrid models have been created using a machine learning classifiers on the CNN feature. The difference in data processing capability needed for patterns based on necessary CNN was Ms. The CNN net had a best accuracy of 98.95 0.9 as precision, recall and F1-score. the planned design increased exactness over audio classification sorters. Also, all three models succeeded in classifying bird audio and proved to be valid. This thesis uses CNN view-based procedures to categorize bird species. The imagination process knows how to consider the entire bird and whereas CNN transmit learning just needs a small experiment from the dataset, passes through with the operation pipes after training, and uses it for classification. Results show that manual feature retrieval and machine learning methods outperform baseline.

REFERENCES

[1] Tianjun Xiao, Yichong Xu, Kuiyuan Yang, Jiaxing Zhang, Yuxin Peng, and Z. Zhang, "The application of two-level attention models in deep convolutional neural network for fine-grained image classification," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2015.

[2] M. Simon and E. Rodner, "Neural Activation Constellations: Unsupervised Part Model Discovery with Convolutional Networks," 2015 IEEE International Conference on Computer Vision (ICCV), Dec. 2015.

[3] Y.; Li, J.; Xiao, H.; Jin, X.; Yan, S.; Feng, J. "Path and Dual Consistency," Constraint Networks, pp. 319–354.

[4] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear CNN Models for Fine-Grained Visual Recognition," 2015 IEEE International Conference on Computer Vision (ICCV), Dec. 2015.

[5] B. Zhao, X. Wu, J. Feng, Q. Peng, and S. Yan, "Diversified Visual Attention Networks for Fine-Grained Object Classification," IEEE Transactions on Multimedia, vol. 19, no. 6, pp. 1245–1256, Jun. 2017.

[6] X. Zhang, H. Xiong, W. Zhou, W. Lin, and Q. Tian, "Picking Deep Filter Responses for Fine-Grained Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016.

[7] Y. Zhang, X.-S. Wei, J. Wu, J. Cai, J. Lu, V.-A. Nguyen, and M. N. Do, "Weakly Supervised Fine-Grained Categorization With Part-Based Image Representation," IEEE Transactions on Image Processing, vol. 25, no. 4, pp. 1713–1725, Apr. 2016.

[8] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016.

[9] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017.

[10] W. Zeng, Y. Wang, and R. Jiang, "Integrating distal and proximal information to predict gene expression via a densely connected convolutional neural network," Jun. 2018.

[11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016.

[13] Y. Bengio "Biophysical Mechanisms of Invariance: Unsupervised Learning, Tuning, and Pooling," Visual Cortex and Deep Networks, 2016

[14] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear CNN Models for Fine-Grained Visual Recognition," 2015 IEEE International Conference on Computer Vision (ICCV), Dec. 2015.

[15] G. Zheng, M. Tan, J. Yu, Q. Wu, and J. Fan, "Fine-grained image recognition via weakly supervised click data guided bilinear CNN model," 2017 IEEE International Conference on Multimedia and Expo (ICME), Jul. 2017.

[16] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a Similarity Metric Discriminatively, with Application to Face Verification," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05).

[17] S. R. K, B. C. K, O. ooruchintala, D. B. Mandru, and kallam suresh, "Deep Convolution Neural Networks Learned Image Classification for Early Cancer Detection Using Lightweight," Oct. 2021.

[18] D. Zhai, X. Liu, H. Chang, Y. Zhen, X. Chen, M. Guo, and W. Gao, "Parametric local multiview hamming distance metric learning," Pattern Recognition, vol. 75, pp. 250–262, Mar. 2018.

[19] E. Hoffer and N. Ailon, "Deep Metric Learning Using Triplet Network," Lecture Notes in Computer Science, pp. 84–92, 2015.

[20] R. Lu, K. Wu, Z. Duan, and C. Zhang, "Deep ranking: Triplet MatchNet for music metric learning," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Mar. 2017.

[21] J. M. Hancock, "Bayesian Network (Belief Network, Causal Network, Knowledge Map, Probabilistic Network)," Dictionary of Bioinformatics and Computational Biology, Oct. 2004.

[22] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning Deep Features for Discriminative Localization," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016.

[23] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, Apr. 2015.

[24] [24] Ahmed, Saadaldeen. (2019). Breast Birds audio detection Detection and Image Evaluation using Augmented Deep Convolution Neural Networks

[25] H. Shen, "Towards a Mathematical Understanding of the Difficulty in Learning with Feedforward Neural Networks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Jun. 2018.

[26] T.-Y. Lin and S. Maji, "Improved Bilinear Pooling with CNNs," Procedings of the British Machine Vision Conference 2017, 2017.

[27] S. Yaghoubi and G. Fainekos, "Hybrid approximate gradient and stochastic descent for falsification of nonlinear systems," 2017 American Control Conference (ACC), May 2017.

[28] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," 2017 IEEE International Conference on Computer Vision (ICCV), Oct. 2017.

[29] M. A. Pedziwiatr, M. Kümmerer, T. S. A. Wallis, M. Bethge, and C. Teufel, "Meaning maps and saliency models based on deep convolutional neural networks are insensitive to image meaning when predicting human fixations," Cognition, vol. 206, p. 104465, Jan. 2021.

[30] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks," 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Mar. 2018.

[31] J. Fu, H. Zheng, and T. Mei, "Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-Grained Image Recognition," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017.

[32] S. R. A. Ahmed and E. Sonuç, "Deepfake detection using rationale-augmented convolutional neural network," Applied Nanoscience, Sep. 2021.

[33] M. T. Mahmood, S. R. A. Ahmed, and M. R. A. Ahmed, "Detection of vehicle with Infrared images in Road Traffic using YOLO computational mechanism," IOP Conference Series: Materials Science and Engineering, vol. 928, no. 2, p. 022027, Nov. 2020.

[34] S. Khalid Abdulateef, S. R. Ahmed AHMED, and M. Dawood Salman, "A Novel Food Image Segmentation Based on Homogeneity Test of K-Means Clustering," IOP Conference Series: Materials Science and Engineering, vol. 928, no. 3, p. 032059, Nov. 2020.

[35] M. R. AHMED, S. R. AHMED, A. D. DURU, O. N. UÇAN, and O. BAYAT, "An Expert System to Predict Eye Disorder Using Deep Convolutional Neural Network," Academic Platform Journal of Engineering and Science, vol. 9, no. 1, pp. 47–52, Jan. 2021.

[36] M. Waleed, A. S. Abdullah, and S. R. Ahmed, "Classification of Vegetative pests for cucumber plants using artificial neural networks," 2020 3rd International Conference on Engineering Technology and its Applications (IICETA), Sep. 2020.

[37] H. Mechria, M. Gouider, and K. Hassine, "Breast Cancer Detection using Deep Convolutional Neural Network," Proceedings of the 11th International Conference on Agents and Artificial Intelligence, 2019.

[38] S. R. A. Ahmed, O. N. UÇAN, A. D. DURU and O. BAYAT, "BREAST CANCER DETECTION AND IMAGE EVALUATION USING AUGMENTED DEEP CONVOLUTIONAL NEURAL NETWORKS", AURUM Mühendislik Sistemleri ve Mimarlik Dergisi, vol. 2, no. 2, pp. 121-129.