Bird Identification Based On the Sound Using Neural Networks

Gowthami Gowda J. S.[1*] and Nalini N.[2 1*]

Computer Science and Engineering, Nitte Meenakshi Institute
of Technology.
[2]Computer Science and Engineering, Nitte Meenakshi Institute of Technology.

Abstract

The paper proposes to identify the bird call using various image and audio classification techniques. The paper brings about a neu- ral network model in combination of machine learning classifiers. To identify the bird call in the audio file by directing the model to pay close attention to the spectrogram images extracted from the audio clippings of birds. The designed set of models achieved the accuracy between the ranges of 89% to 98% on the testing audio set.

Keywords: ANN, ANN+SVC, ANN+KNC, LSTM, LSTM+SVC, LSTM+KNC, Audio classification

1 Introduction

1.1 Background

Over the years, the audio recognition has savored the fruit of a progressively increasing interest among the individuals [1, 2]. The commerciality of DL and its many varieties of layered-network models have played a key role in helping to solve the ML problems of regression and classification effortlessly. Of all the myriad categories present in the audio varieties; the cornerstone of the project is bird call-sounds only. The earmark audience are tone deaf people, nature lovers/explorers, bird lovers and bird sanctuary tourists; helping them to tell between bird species just by its audible sounds.

1.2 Brief History

The social beings, humans are always interested in monitoring their surrounding environment using technology; which seems to have increased during recent years. The humans are of a curious mind, always trying to communicate and understand other species like dolphins, dogs, whales; The presented project is also one such communication with other species widely called birds; especially in identifying their species just from their sounds has scooped and put as-it-is in much scrutiny. Whatever might be the case, it is not to be neglected that there are various number of bird species; more to be discovered; and even more categories to be studied before analyzing and recognition of the correct and precise bird species; but it’s way more gruesome than one can comprehend with so little statistics get-at-able on such a amazing creature apportion this orb with us.

A diverse number of parameters (tweet/twitter/cheep/chirp of the bird are considered) are to be profiled-out for classification speshly here; where pre- dominance of it is is on how to extract the information from unstructured data, namely audio recordings of bird call; working mainly focusing on MFCC filters [1–4], wavelets and LPC coefficients [5, 6]. Lee et al. [7], seems to have used a similar contradictory approach for the image (spectrogram) characteristics.

1.3 Applications

The proposal for the project was basically made to satisfy its unique requirement and irreplaceable acceptance in the real-world. Some list as;

1. Music industries:- Musicians tend to be more responsive to sound than others; natural sounds bring them bliss and are obviously more inquisitive in birds; even more so that they go to lengths to even include it in the composition.
2. Tourists:- Excursionists nurture a mind of marvel and reckon they ought to know the sounds they deduct during their excursion.
3. Green Panther:- Activists and nature lovers work towards protecting environmental habitats; being interested in nature; includes bird sound also.
4. Therapy:- It is trending to use birds and their sound to give physiotherapy for the mentally weaker individuals; some bird sounds are used to treat some patients, and some others; the patients may also be interested in knowing the kind of birds they are receiving treatment from.
5. Child growth:- Birds or any pets for that matter is supportive in child’s blooming; hearing and learning to identify the birds are one such develop- ment; equivalent to upskilling them to identify animals visually, it is also promising to teach them to recognize birds from their sounds.
6. Identify danger:- If encountered with the sound of predatory birds, it is best advised to avoid them; so best to identify the birdbeforespottingthem.

7. Find direction:- If blessed enough to identify the migratory birds when you are lost on your way; birds are best friends at the time.
8. Tone-deaf people (Amusia):- the people with amusia fail to distinguish between different sounds as it all sounds the same to them. So, this research work helps them in satisfying their introverted minds by helping them classify the bird based on the bird sound they encounter.

1.4 Research Motivation

Bird species identification based on song/vocals is an application where the data in the form of audio file (having .wav file), may contain a lot of noise which becomes a great hurdle in recognizing the bird type correctly. As the project is being carried out keeping in mind the tone-deaf (amusia) individuals; who can't pick out differences in pitch or follow the simplest tunes and all sound the same to them. This research work helps them to some extent to solve one such problem of helping them identify the bird sound they tend to hear; record it and identify the bird.

1.5 Problem Statement

Birds of the same species sound differently from one region to another, maybe even across countries due to migration, weather conditions and so on, which affect the bird's sound. Also, hearing the bird and identifying their kind is rather an impossible task for tone-deaf people. This research project work is aimed at those rarely occurring tone-deaf individuals.

2 Literature Survey

2.1 Related Work

Even when excluding Xeno-Canto database, there are assorted numbers of datasets publicly made available. Such as NIPS-4B dataset [15], Warblr dataset [19], Birdbox-Full-Night dataset [14], Poland-NFC dataset [17], Freefield-1010 dataset [18], and Chernobyl dataset [10]. Over the last few years, various techniques have been researched upon. An approach on the random forests along with RNN-LSTM [11–13]; training lesser statistical features of spectrograms and also the same of audio files. Never- theless, when compared with traditional model-DCNNs, surprisingly, DCNN's still turned out to be the best approach.An ANN based model using Power Spectral Density (PSD) method [1] for preprocessing of audio data. PSD is inordinately worthwhile for extricate temporal measurements of periods of sound and somewhat occupancy of hush within songs. Once the temporal complement of spectrograms is cooked, it is then utilized to scrutinize song patterns and syllables in birdsong.Extending the supremacy on the spike based model [2]; in which case it bestows itself with the affair of linking the bird species built upon vocalization or call. Also used matlab for trickling and probing the audio format data scrutinizing that they have solely .wav format. Contemplating on features such as sonorous, entropy, tonality, timbre, loudness, etc., to associate the birds steer a SNN with permutation-frequency matrix [2].Drudging on pin-down the species of birds based on its vocals by mash-up of MFCC, average spectra and noise robust gauging [3]. Use of audio-signals processing techniques to unsheathe nearly 35 features from bird song; which are then narrowed down using Linear Discriminant Analysis (LDA); before feeding the Nearest Centroid (NC) classifiers to do the job; found to transcend compound classifiers, specifically SVM and Adaboost classifiers, by attaining an approximate 96% accuracy [3].Embodying use of an unsupervised algorithm in duplet instant. Foremost, a spectrogram enhancement technique is advocated to employ Savitzky-Golay (MWSG) filters [4]. Furthermore, milking the exceptionally structured Time-Frequency (T-F) cues in discrete leadership from enriched spectrograms discernment [4]. Also, bagging frame-level binary decisions of sound from unidirectional spectrograms to make closing decree. Anatomization the NCD merit [5], use bird vocals to assort the bird species as an evaluation metrics to the bird species pinpointing via audio instances. To begin with, the practitioners [5] have appraised the repercussions of conflict- ing compression tactics concerning the bird audio illustrated from the datum apiece whose performance is pondered by appertaining the distance matrix by projection mapping and hierarchical clustering and both are metered. Also, challenged novel deep learning models to promote their commensu- rately sound information. Also, a cGAN based model with auxiliary classifiers, exhibiting sound coders to uproot preferable feature delineation from audio tape-recordings and then conjure spectrogram images analogous to their calls/vocals [6]. A renowned feature descriptor [7] is designed; elementary units are the fixed-duration audio clips. Partition the input audio into equal sized windows, such as 3 or 5 second segments, and then manipulated into the requirement of audio classification. They [7] have made use of 28 species with a test accuracy of 86.30% for 3 second segments and 94.62% for 5 second segments. Anatomizing the results [7], it appears that preponderant bird species are correctly discerned maneuver ART descriptors, contradicting some results which showed soaring classification bloomers. Demonstrating an improvised spectrogram [8], which works towards increasing the performance of the applied algorithm (using Markov-renewal process); built on the distribution-derivative method, trail down sound- changeables patterns? It is also a necessity to address the task using a pre-trained model, VGG- 16 model, on mel-spectrogram. Also, use probability scores to verify if the audio clipping/recording actually has bird-sound in them (maximum probabil- ity scores) [9]. Feed mel-spectrograms as an input for VGG-16 model and do a comparative study on its performance with that of the MFCC-DNNmodel. After going to such extremes, it can be seen that the effort was not in vain, as the proposed [9] model outperforms the MFCC-DNN model.

3 Proposed System

3.1 Project Approach

The project approach diverged towards implementing a neural network system to train and classify the bird species. Based on the previously done work [19], it is safe to say that deep neural networks give best results when dealing with unstructured data such as audio data and that too the bird datasets where people are usually not able to distinguish between the two similar sounding birds call. When talking about dataset collections, the audio files are collected from xeno-canto website in mp3 format and later converted into required .wav format. Since the data is now ready the audio data is explored for getting familiar to the data. Data set is again converted into an image dataset by extracting the spectrogram images of each of the audio and later the images are converted into a numerical dataset. Now, to build the model on the prepared dataset I have chosen the neural networks, ANN and LSTM. Neural networks are trained on these numerical data and at the end of the terminal dense layer the feature extracted is fed into the machine learning classifiers to classify the bird call and identify the species of the bird. Finally, since all the training and testing of the model is done; the model is subjected to a series of performance evaluations by testing them on the known and unknown data and their values are recorded. The accuracy at which the model has been trained and tested is compared between all the proposed models and the same has been plotted.

3.2 Proposed System

The presented work/system is a combination of neural networks and machine learning classifiers. The input to the system is the audio file with format, .wav format; only accepts .wav extension file. Each of these has to undergo a series of audio processing to output the correct species of bird. The models proposed are discussed further in the report. To get an insight on how the proposed model looks like, it can be summed up in Fig. 1.

3.3 Steps for Bird Species Classification

1. Firstly, the data collected are in the form of .wav format; which later have to be converted into images by spectrogram extraction from all this audio and finally these images are converted into metadata or numerical data to work on.
2. The audio data is converted into image; spectrogram image using nonuni- form fast Fourier transform (NFFT); and storedinthefiletobeusedlater.

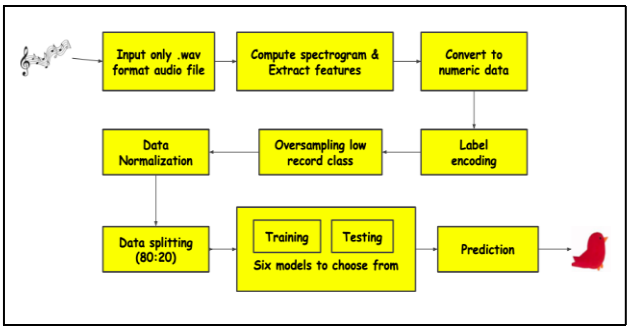

Input only .wav format audio file
Compute spectrogram & Extract features
Convert to numeric data
Data Normalization
Oversampling low record class
Label encoding
Data splitting (80:20)
Training
Testing
Six models to choose from
Prediction

Fig. 1: Proposed System

3. Converting the spectrograms into a Numeric data file.
4. Every label, here 7 is mapped to numeric from 0 to 6.
5. Since there are only about 30 to 32 datasets for every bird species; it isn't sufficient to work with neural networks. So oversample the data to 200 datasets for every species of bird present.
6. Data has to be normalized; data type in the dataset may be of any type, so normalization is done.
7. Data split in 30:70 ratios; for testing and training.
8. Every neural network is trained for 20 epochs subjected to Adam optimizer and sparse cross entropy with .0003 learning rate.
9. The proposed model after undergoing all the aforementioned steps has to be saved for further use.
10. The trained model is evaluated.

4 Implementation

4.1 Data Modeling

Various models are built on the feature spectrogram image generated from the .wav file of each bird call and comparison is done. Initially, a simple ANN and LSTM model are seriatimly built and then each of its extracted features from the dense layer is fed into the machine learning algorithms such as SVC and KNC, respectively.

4.1.1 Model-1: ANN

The model of Artificial Neural Network (ANN) or most commonly called Con- volutional Neural Network (CNN) taken as a substructure for Model-1 is as portrayed in Fig. 2. A designed ANN architecture of Model-1, comprises con- nected layers. The sequential ANN model has three alternative dense layers with units 128 and activation function as ReLU; and dropout layer of size m0.2 to reduce over fitting the network; third layer with 32 units. Lastly, the last layer is of units 7, which is the expected number of output labels using soft max activation function. Adam optimizer compiles the model along with; categorical cross entropy, learning rate as in Fig. 2.

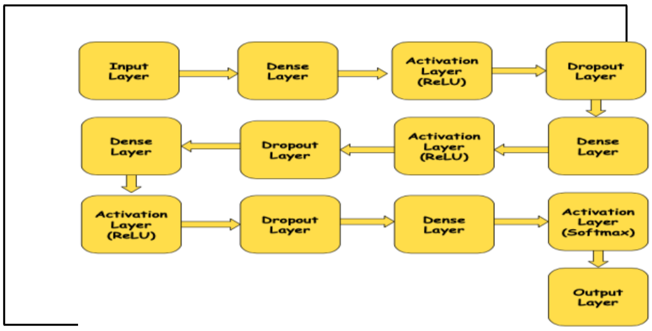
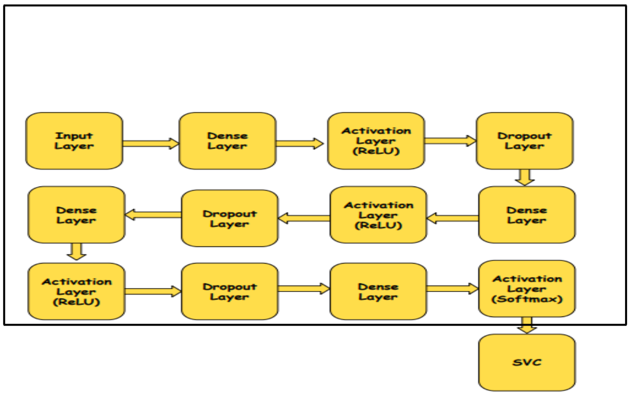


Fig. 2: Model-1: ANN architecture

4.1.2 Model-2: ANN+SVC

As part of Model-2, the ANN model is combined with SVC. The feature extracted from the terminal dense layer of Model-1 as designed in Fig. 2, is fed to the SVC. The resultant architecture turns out to be Fig. 3. This architecture is mainly proposed because the ANN can be best used for feature extraction and SVC correctly classifies the data. Fig. 3 shows the SVC implementation i.e the feature extracted from ANN is fed into SVC for classification and inturn increases the overall Model-2 accuracy.



4.1.3 Model-3: ANN+KNC

Akin to Model-2; Model-3 also uses a classifier called KNC to classify the feature obtained from the ANN, Fig. 4.

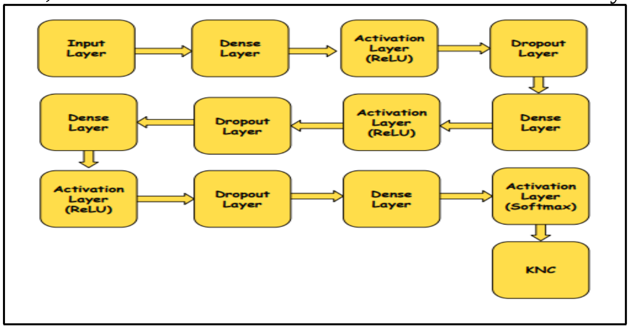
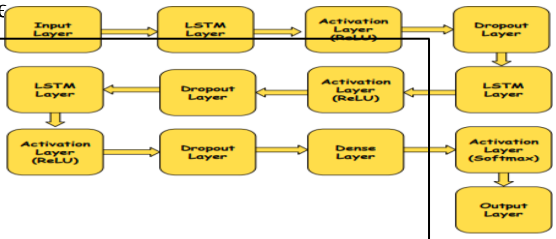


Fig. 4: Model-3: ANN+KNC architecture

4.1.4 Model-4: LSTM

The designed LSTM architecture of Model-5, comprises connected layers. The sequential LSTM model has three alternative dense layers with units 128 and activation function as ReLU; and dropout layer of size 0.2 to reduce overfit- ting the network; third layer with 32 units. Lastly, the last layer is of units 7, which is the expected number of output labels using softmax activation func- tion. Adam optimizer compiles the model along with; categorical cross entropy, learning rate as in Fig. 5.

Fig. 5: Model-4: LSTM architecture



4.1.5 Model-5: LSTM+SVC

As part of Model-5, the LSTM model is combined with SVC. The feature extracted from the rear most dense layer of Model-4 as coined in Fig. 5, is fed to the SVC. The resultant architecture turns out to be Fig. 6. This architect- true is predominantly proposed because the LSTM can be best used for feature extraction and SVC correctly classifies the data. Fig. 6 shows the SVC imple- mentation i.e the feature extracted from LSTM is fed into SVC for classification and intern increases the overall Model-6accuracy.

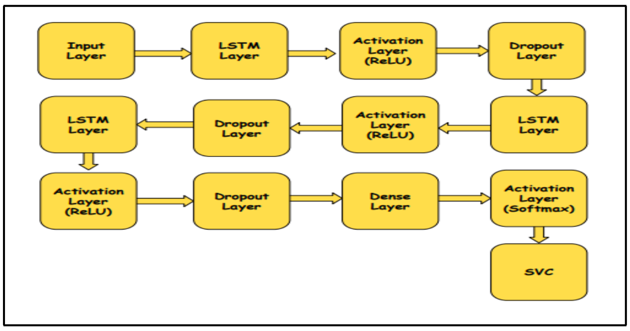
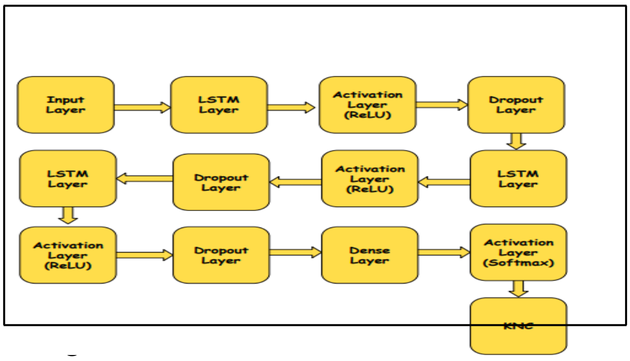


Fig. 6: Model-5: LSTM+SVC architecture

4.1.6 Model-6: LSTM+KNC

Akin to Model-5; Model-6 also uses a classifier called KNC classifier to classify the feature obtained from the LSTM, Fig. 7.



5 Result and Analysis

5.1 Performance of ANN and LSTM

Fig. 8 and Fig. 9 show the performance of the ANN and LSTM in terms of accuracy and identically Fig. 10 and Fig. 11 for the loss, respectively. Each of these models were trained for epoch 20 each after which no increase in accuracy or decrement in loss were to be found, so they were trained on epoch 20 to get best-fit performance of the model. The model is subjected to 7 species of birds each of 200 datasets, and they incurred the accuracy ranging from 89% to 98% on the test data where the split oftest-trainwasin1:4ratio.

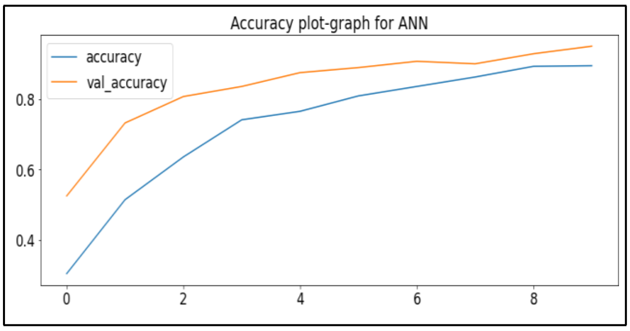
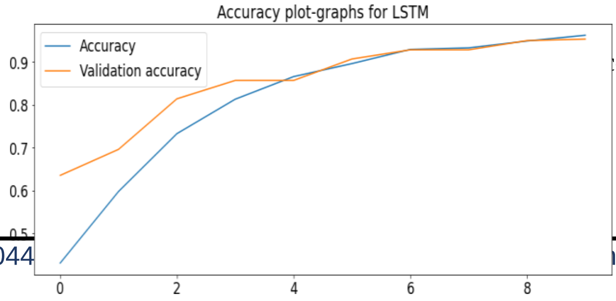


Fig. 8: ANN Accuracy



curacy

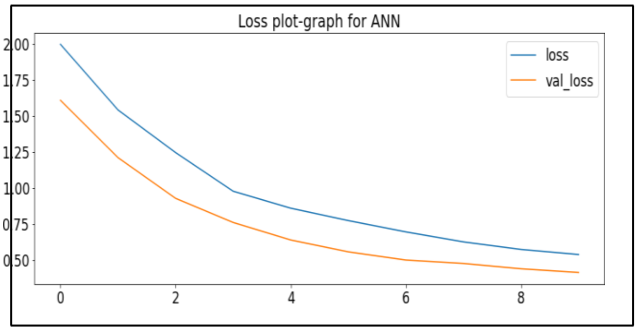


Fig. 10: ANN Loss

5.2 Accuracy Score

Accuracy score is one of the metrics used in result analysis of the trained neural networks models. The accuracy score tells about how accurate the trained model performs prediction on the test data. Fig. 12 is the accuracy score of the all six models trainedaspartofthisproject.

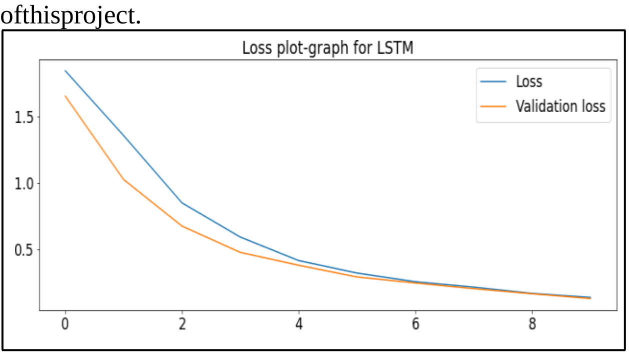
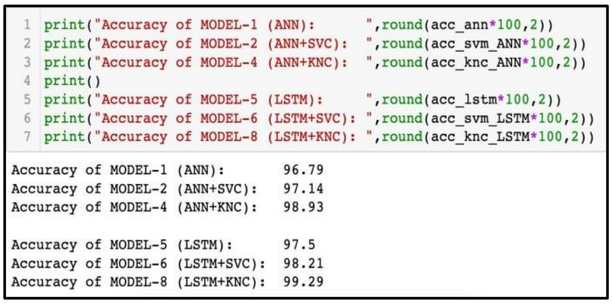


Fig. 11: LSTM Loss



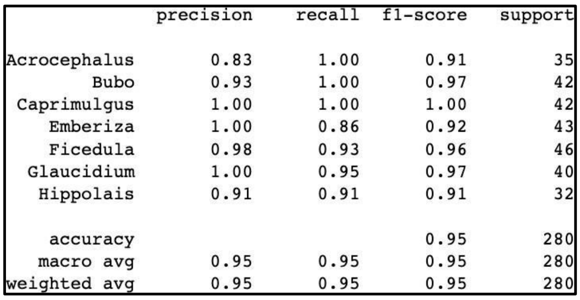
```
print("Accuracy of MODEL-1 (ANN):      ",round(acc_ann*100,2))
print("Accuracy of MODEL-2 (ANN+SVC):  ",round(acc_svm_ANN*100,2))
print("Accuracy of MODEL-4 (ANN+KNC):  ",round(acc_knc_ANN*100,2))
print()
print("Accuracy of MODEL-5 (LSTM):     ",round(acc_lstm*100,2))
print("Accuracy of MODEL-6 (LSTM+SVC): ",round(acc_svm_LSTM*100,2))
print("Accuracy of MODEL-8 (LSTM+KNC): ",round(acc_knc_LSTM*100,2))
```

```
Accuracy of MODEL-1 (ANN):       96.79
Accuracy of MODEL-2 (ANN+SVC):   97.14
Accuracy of MODEL-4 (ANN+KNC):   98.93

Accuracy of MODEL-5 (LSTM):      97.5
Accuracy of MODEL-6 (LSTM+SVC):  98.21
Accuracy of MODEL-8 (LSTM+KNC):  99.29
```

Fig. 12: Accuracy Score

5.3 Classification Report

Fig. 13 shows the Classification Report of the Model-1; ANN. Classification Report of the Model-1 gives an insight about the precision, recall, F1 score and support for each of the bird species present in the dataset.

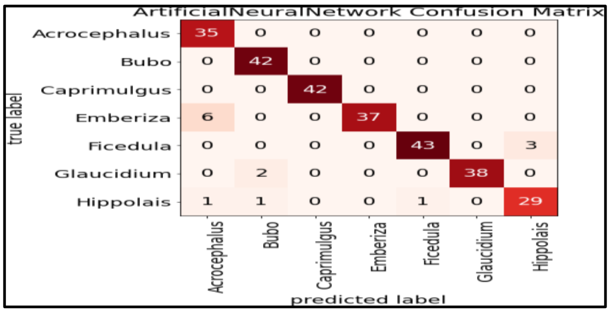


```
              precision    recall  f1-score   support

Acrocephalus       0.83      1.00      0.91        35
        Bubo       0.93      1.00      0.97        42
 Caprimulgus       1.00      1.00      1.00        42
    Emberiza       1.00      0.86      0.92        43
    Ficedula       0.98      0.93      0.96        46
  Glaucidium       1.00      0.95      0.97        40
   Hippolais       0.91      0.91      0.91        32

    accuracy                           0.95       280
   macro avg       0.95      0.95      0.95       280
weighted avg       0.95      0.95      0.95       280
```

Fig. 13: ANN Classification Report

5.4 Confusion Matrix

Confusion Matrix is often used to tabulate the performance of the models based on a set of test data whose values are known. Fig. 14 is the confusion matrix for the Model-1 (ANN)



	Acrocephalus	Bubo	Caprimulgus	Emberiza	Ficedula	Glaucidium	Hippolais
Acrocephalus	35	0	0	0	0	0	0
Bubo	0	42	0	0	0	0	0
Caprimulgus	0	0	42	0	0	0	0
Emberiza	6	0	0	37	0	0	0
Ficedula	0	0	0	0	43	0	3
Glaucidium	0	2	0	0	0	38	0
Hippolais	1	1	0	0	1	0	29

Fig. 14: ANN Confusion Matrix

5.5 Comparing Models

Fig. 15 shows the overall model accuracy in the form of bar plot. It can be noted that the Model-1 gives the least accuracy and Model-6 gives the highest accuracy.

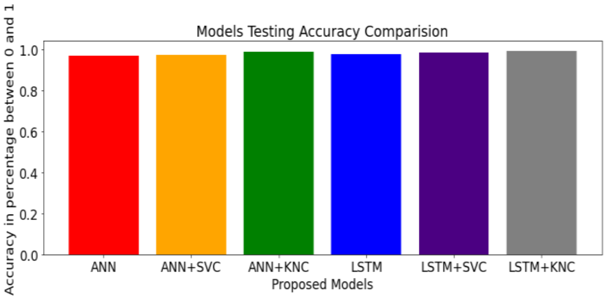


Fig. 15: Comparing Models

Conclusion

The paper has been proposed to identify the birds species correctly by training them on neural networks and suitable classifiers to avoid any accuracy defaults. The goal to present a system that is dependable and accurately classify and

Identify calls to help tone-deaf individuals is achieved. The system is malleable enough; the tone-deaf people can easily make use of it without any dubiousness. The microphone plays an important role in providing the inputs to the system.

Future work and way forward

The main limitations faced is the low computational power on the local system, so moving forward the system to be deployed on the cloud; hence the load on the local desktop can be reduced significantly. Since this paper's main focus was on one-on-one mapping for bird and audio file, in the future work can be carried upon multiple bird sound classification.

References

[1] N. R. Koluguri, G. N. Meenakshi and P. K. Ghosh, ”Spectrogram Enhancement Using Multiple Window Savitzky-Golay (MWSG) Filter for Robust Bird Sound Detection,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 6, pp. 1183-1192, June 2019, doi: 10.1109/TASLP.2017.2690562.

[2] R. Mohanty, B. K. Mallik, and S. S. Solanki, “Automatic bird species recognition system using neural network based on spike,” Applied Acous- tics, vol. 161, 2020.

[3] M. Ramashini, P. E. Abas, U. Grafe and L. C. De Silva, ”Bird Sounds Classification Using Linear Discriminant Analysis,” 2019 4th International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), 2019, pp. 1-6, doi: 10.1109/ICRAIE47735.2019.9037645.

[4] N. R. Koluguri, G. N. Meenakshi and P. K. Ghosh, ”Spectrogram Enhancement Using Multiple Window Savitzky-Golay (MWSG) Filter for Robust Bird Sound Detection,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 6, pp. 1183-1192, June 2019, doi: 10.1109/TASLP.2017.2690562.

[5] G. Sarasa, A. Granados and F. B. Rodriguez, ”An approach of algo- rithmic clustering based on string compression to identify bird songs species in xeno-canto database,” 2019 3rd International Conference on Frontiers of Signal Processing (ICFSP), 2019, pp. 101-104, doi: 10.1109/ICFSP.2017.8097150.

[6] J. Y. Shim, J. Kim and J. -K. Kim, ”S2I-Bird: Sound-to-Image Generation of Bird Species using Generative Adversarial Networks,” 2020 25th Inter- national Conference on Pattern Recognition (ICPR), 2021, pp. 2226-2232, doi: 10.1109/ICPR48806.2021.9412721.

[7] C Lee, S Hsu, J Shih, and C Chou, “Continuous birdsong recognition using gaussian mixture modeling of image shape features,” IEEE Transactions on Multimedia, vol. 15, no. 2, pp. 454 – 464, 2013.

[8] Dan Stowell, Saˇso Muˇseviˇc, Jordi Bonada and Mark D. Plumbley, “Improved multiple birdsong tracking with distribution derivative method and Markov renewal process clustering” in arXiv:1302.3462v2 [cs.SD], https://doi.org/10.1109/ICASSP.2013.6637691.

[9] R. Rajan and N. A, ”Multi-label Bird Species Classification Using Trans- fer Learning,” 2021 International Conference on Communication, Control and Information Sciences (ICCISc), 2021, pp. 1-5, doi: 10.1109/IC-CISc52257.2021.9484858.

[10] Paul Kendrick et al. Bird Vocalisation Activity (BiVA) database: anno- tated soundscapes from the Chernobyl Exclusion Zone. Sept. 10, 2018.

[11] Mario Lasseck. ”Bird Song Classification in Field Recordings: Winning Solution for NIPS4B 2013 Competition”. In: Proceedings of the Workshop on Neural Information Processing Scaled for Bioinformatics (Lake Tahoe, USA). Ed. by Herv´e Glotin et al. Jan. 2013, pp. 176–181.

[12] Mario Lasseck. ”Improved Automatic Bird Identification through Decision Tree based Feature Selection and Bagging”. In: Working Notes of CLEF 2015 - Conference and Labs of the Evaluation Forum (Toulouse, France). Ed. by Linda Cappellato et al. Vol. 1391. CEUR Workshop Proceedings. CEUR, Sept. 2015, pp. 1–12.

[13] Mario Lasseck. ”Large-Scale Identification of Birds in Audio Recordings”. In: Working Notes of CLEF 2014 - Conference and Labs of the Evaluation Forum (Sheffield, United Kingdom). Ed. by Linda Cappellato et al. Vol. 1180. CEUR Workshop Proceedings. CEUR, Sept. 2014, pp. 643-653.

[14] Vincent Lostanlen et al. ”Birdbox-Full-Night: A Dataset and Benchmark for Avian Flight Call Detection”. In: International Conference on Acous- tics, Speech and Signal Processing - ICASSP 2018 (Calgary, Canada). IEEE Computer Society, 2018, pp. 266–270. ISBN: 978-1-5386-4658-8. DOI: 10.1109/ICASSP.2018.8461410.

[15] Veronica Morfi et al. ”NIPS4Bplus: a richly annotated birdsong audio dataset.” In: PeerJ Computer Science 5.e223 (Oct. 7, 2019), pp. 1–12. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.223.

[16] Lukas Mu¨ller and Mario Marti. ”Bird sound classification using a bidirec- tional LSTM”. In: Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum (Avignon, France). Ed. by Linda Cappellato et al. Vol. 2125. CEUR Workshop Proceedings. CEUR, Sept. 2018, pp. 1–13.

[17] Hanna Pamu-la et al. ”Adaptation of deep learning methods to nocturnal bird audio monitoring”. In: Archives of Acoustics 42.3 (Sept. 30, 2017), p. 149-158. ISSN: 0137-5075.

[18] Dan Stowell and Mark D. Plumbley. ”An open dataset for research on audio field recording archives: freefield1010”. In: CoRR 1309.5275 (Oct. 1, 2013), pp. 1–10.

[19] Dan Stowell et al. ”Automatic acoustic detection of birds through deep learning: The first Bird Audio Detection challenge”. In: Methods in Ecol- ogy and Evolution 10.3 (Mar. 2019), pp. 368–380. ISSN: 2041-210X. DOI: 10.1111/2041-210X.13103.