What Did the Chicken Say?: A Multi-class Classification Method on Chicken Vocalisations

Kumar Mangalam[1], Soumya Sarkar[2], Yakul Dogra[3], Mukesh Saini[4], Neeraj Goel[5]

[1] Indian Institute of Technology, Ropar
{2022aim1002,2022csm1013,2022csm1019}@iitrpr.ac.in
[2] Indian Institute of Technology, Ropar
mukesh@iitrpr.ac.in
[3] Indian Institute of Technology, Ropar
neeraj@iitrpr.ac.in

Abstract. Poultry is one of the largest industries in India, with a market size of USD 23.07 billion in 2022. It is imperative that the health conditions of the poultry chickens are monitored to ensure optimum output. However, due to the size of poultry farms it is a difficult task to keep track of individual chickens and ensure healthy conditions. Our current proposal aims to target this very problem statement using chicken vocalizations. We use various machine learning models to process audio signals of chicken vocalisations and try to find the best implementation combinations to get the best classification results. We find that our method gives **92.23%** accuracy and shows potential in real-life applications.

Keywords: Chicken distress, Audio processing, Agriculture, Deep Neural Networks

1 Introduction

Indian poultry industry reached a market size of INR 1905.3 Billion in 2022 and is projected to reach INR 3477.8 billion by 2028 [4]. As demand keeps increasing, so does the need to not just expand the poultry farms, but ensure the healthiest of conditions in the broilers so that mortality rates are low. To achieve this, there is an increasing demand for monitoring the health of the livestock. In this aspect, the use of non-invasive methods is the way to go for an entirely automated and remote farming setup. One such methodology by Mcloughlin [9] is based on bioacoustic methods and is becoming increasingly popular.

Bioacoustics is the study of the production, transmission and reception of animal sounds. This includes not only the vocalizations of animals such as birds and mammals, but also the sounds that can be produced by insects. Methods in bioacoustics are becoming increasingly automated, with researchers deploying autonomous recorders that are capable of automatically collecting data. Animal welfare assessment and monitoring could benefit from increased use of such automated methods.

There has been slow and steady research in this field as applied on poultry, and results have been showing promise. Caroll et al. [1] proposed a C4.5 classifier where they trained the C4.5 decision tree classifier using the histogram of clustered MFCCs, with this they achieved an accuracy of 73.4% and were one of the earliest papers exploring this avenue. However, it has been seen that accuracy of classical methods are not as high as deep learning-based methods [9]. In that sense, it makes sense to explore avenues based on neural-network methods.

Chicken distress calls can be identified through their vocalizations using various methods of acoustic analysis. Chickens emit different vocalizations depending on their emotional state and environmental conditions, including distress calls when they are experiencing pain, fear, or discomfort. One method of identifying chicken distress calls is by analyzing their Mel Spectrograms [9]. By analyzing their Mel Spectrograms, it is possible to identify patterns and features that are specific to these vocalizations.

In this paper we employ a deep-learning methodology to classify chicken vocalisations into three categories - distress (more specifically, distress due to visitors), feeding vocalisations and egg-laying vocalisations. We show that even with standard phone recording we achieve promising results in terms of accuracy, recall, and precision - thereby opening doors to a practical application on a day-to-day basis. We compare between a regular CNN model and a transfer-learning based VGG16 model on several key paramaters. We also show what is the ideal value of split that needs to be carried out on any audio sample to achieve the best results. In our knowledge, we are the first to carry out this kind of analysis - there is only one paper by Thomas et al. [12] which classifies chicken audio data into multiple classes, and that too was based on a different set of classes with audio being collected under laboratory conditions. Thus, having no comparable baseline, we provide analysis based on our model choices only.

2 Related Work

One of the major origins of health monitoring came from cattle health monitoring [7]. One recent application to poultry by Raj et al.[10] was done in 2018, pairing it with IoT to create an end-to-end poultry management systems. It uses various components such as gas sensors, microphones, thermal humidity sensors, RGB cameras, and a thermal camera. However, it must be noted that this system would be quite expensive to implement, which is why we were drawn to the most cost-effective methods to achieve comparable results, thereby zeroing in on detection using just sound. The paper [9] describes several methods based on bioacoustics that have been recently developed to monitor animals.

An earlier paper by Curtin et al.[3] tries to incorporate bioacoustic analysis to predict chicken stress. Their basis for stress was to count the number of vocalisations using inexpensive audio setups to detect stress. Another paper that worked with stress detection was by Huevel et al.[13], where they use of thermographic imaging and microphones in obtaining objective indicators for acute stress in laying hens.

Perhaps the work that was closest to ours and which forms the base of our research was the paper by Mao et al. [6]. It collected data from production facilities based in China from broilers where chickens were kept in stacked cages (three per stack). They specified the position of their recording device - a Zoom H4n Pro, which retails at upwards of Rs. 20000 in India, making this audio collection philosophy less suited to real-world audio recording situations. They annotated the data using software and carried out initial prepossessing, dividing the audio chunks into non-overlapping segments of 1s each. They then converted these audio segments into log-Mel spectograms and classified them on the basis of several models, out of which the best performing one was the VGG11 model. We take a lot of our inspiration from this work, and try to apply techniques to get better generalizability and usefulness in the real world.

In 2023, there was a publication by Thomas et al. [12] which had a very similar problem statement to ours but their approach to solving it was slightly different. Their audio data collection strategy is also, like [6], based on isolated laboratory settings. It was classified into the different classes by first identifying interesting broiler vocalisations using pure-convolution-based pre-trained audio neural networks (PANNs), followed by auditive evaluation and visual inspection of the linear-frequency spectrogram to manually label each sample. The samples were then classified into four distinct types: distress calls, short peeps, warbles and pleasure notes. In cases where none of these vocalizations were detected, the sample was labeled as ”other sound”. They base their model on a simple CNN network and achieve a broiler vocalization classification balanced accuracy of 87.9%.

In related work, there has been exciting research going on in the related domain of disease detection among poultry. One emerging direction we think has the most potential is in the detection of avian flu using chicken vocalizations, as done by Cuan et al. in [2]. it must be noted, however, that in this paper too they have used laboratory conditions for their data. It is an avenue that we might explore in a future paper using data processing techniques and methodologies based on this paper.

3 Methodology

After the input data is provided, using signal processing techniques we generate Mel Spectograms, which we feed to our model to generate the output classification. To train this model we share our data collection and data labelling methodologies below. The overall architecture of the proposed framework is shown in Figure 1.

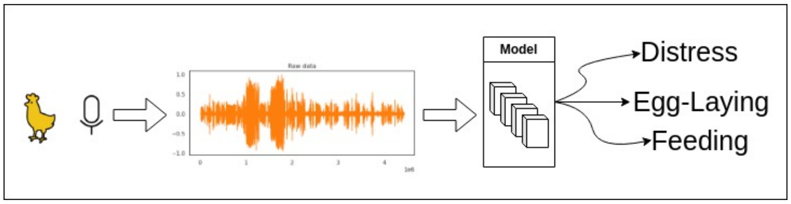

Fig. 1: Overall architecture of the proposed framework

3.1 Audio Collection

Our audio collection philosophy was based on trying to emulate the hustle and bustle of a standard Indian broiler in our audio data so that we are able to process the data and achieve good generalization. To do this, we had to collect the data ourselves, since there were no good sources of data available online. We went to a poultry farm in Bainsa, Shaheed Bhagat Singh Nagar, Punjab to collect our audio. The chickens were placed in tightly-packed cages (four chickens per cage) and were of the common Indian poultry breed (Busra breed). Since it was their egg-laying time, we were able to collect the data for egg-laying and visitor distress. We also later were able to record video footage during their feeding time. We were, overall, able to collect data of size 1 hour approximately using our handheld recording devices (our personal phones - Samsung Galaxy A23, Poco M4 Pro 4G). We collected the data by recording it in video format so that we could later use it to manually annotate. The audio consisted of not just noises from multiple chickens all at once but also background noises such as people talking (very few instances of this exist). The layout of two of the rows (there were four in the single room) has been explained in the Figure 2 below.

Fig. 2: Layout of Broiler We Visited

Table 1: Data Distribution in terms of Mel Spectograms

Split	Dataset	Distress	Egg-Laying	Feeding	Total
1	Train	85	390	1003	1478
1	Validate	12	56	143	211
1	Test	24	111	286	421
0.5	Train	169	779	2004	2952
0.5	Validate	24	111	286	421
0.5	Test	49	224	574	847

3.2 Dataset Annotation

To annotate the data, we had to manually run through all the video and crop out the parts belonging to different classes. The audio was taken from the video footage shot on the farm by first extracting the audio using an online extractor at 128Kbps, and then finding instances of distress in the continuous audio. Thereafter we manually split it without overlap in 0.5s and 1s intervals. The dataset distribution for our intended three classes are given in Table 1.

3.3 Audio Preprocessing and Generating Mel Spectograms

The recordings are then gated using a minimum decibel limit mentioned in Table 2, and then transformed into Mel Spectrograms, which are two-dimensional representations of sound frequencies over time, using the librosa library of functions [8]. This step is important for feature extraction, which is a crucial part of the machine learning process [5]. A sample Mel spectogram is shown here in Figure 3.

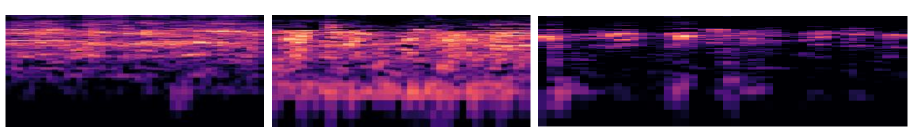

Fig. 3: Mel Spectograms for (left to right): (a) Feeding (b) Egg-Laying (c) Distress

We see the differing nature of the three calls in Figure 3. The x-axis refers to the time duration (1s here). The y-axis denotes the frequency value of the signal. The different colors represent the decibel levels (the darker the color, the louder the sound). Thus we see that egg-laying is relatively quite a calm call, whereas feeding has a lot more activity. However loudness sharply increases during distress, as shown by the large dark spots.

3.4 Classification by Model

The Mel Spectogram was then processed via several different categories of models. In our case we wanted to compare with three classes of models. Our first model is a simple custom neural network model. The architecture of the model is defined as below in Figure 4. For this architecture we end up with less than 300,000 paramaters. In our case we found that it only had a size of 3.4MB. Thus, this is a light and easily deployable model.

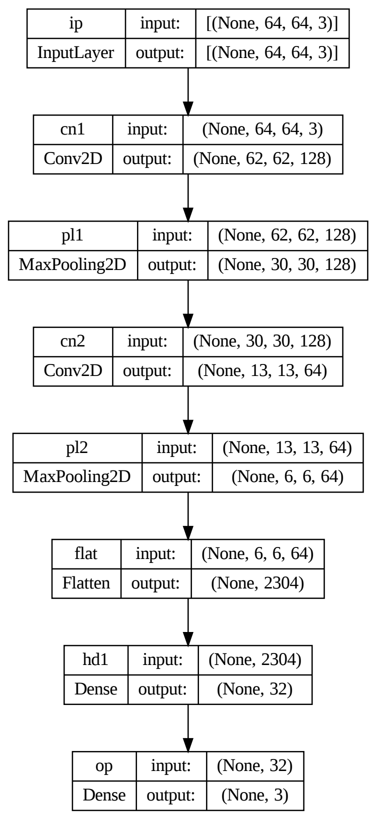

Fig. 4: Model Architectures for Custom CNN

Table 2: Table of Hyperparameters

Paramater	Value
Hidden Layer Activation Function	ReLU
Output Layer Activation Function	Softmax
Early Stopping Minimum Delta	0.001 (for Validation Loss)
Batch Size	32
Librosa Trim Top Decibel	20
Sampling Rate	22050
Number of Mels	128*2

Our second model is a VGG16-based model [11], which was released in 2014, thereby providing a more modern but more complex approach to solving this issue. It involves a pre-trained VGG16 model on which we fine-tuned the VGG16 model with our proprietary dataset. This choice of model is particularly important since VGG16 has been trained on 1,000 different classes, so we felt that it would have been best suited to our problem statement.

4 Results

We run our experiments keeping in mind the various possible combination of choices we have. Our choice of hyperparamaters are given in Table 2. Our base model ran for 32 epochs, while the VGG model ran for 11 epochs, with early stopping for both. For each model we have, we test for both a 0.5s split (Figures 5) and a 1s split Figures 6. We find we achieve a best case categorical accuracy of 92.23% using our base CNN model and split of 1s.

As we can see in Table 3, we achieve a greater accuracy with the base CNN model as compared to the VGG16 model. Whereas we get an accuracy of 92.23% with our base CNN model ,we end up with an accuracy of 90.82% with our VGG16 model. We also achieve a higher precision and recall. We also see that validation error is slightly higher - we manage to achieve an accuracy of 90% for our base model, v/s the VGG16 model which achieves an accuracy of 86%. This perhaps can be explained by the fact that our model is relatively much simpler and hence has less requirement of data to predict more effectively. Another reason is that VGG16 has been trained on a much more generalised dataset which does not consist of specialised images like Mel Spectograms, whereas our model has been specialised to work on Mel Spectograms. That being said, with more data we definitely should see an improvement for the VGG16 model.

We also see that we achieve better results for a split of 1s, although the val-

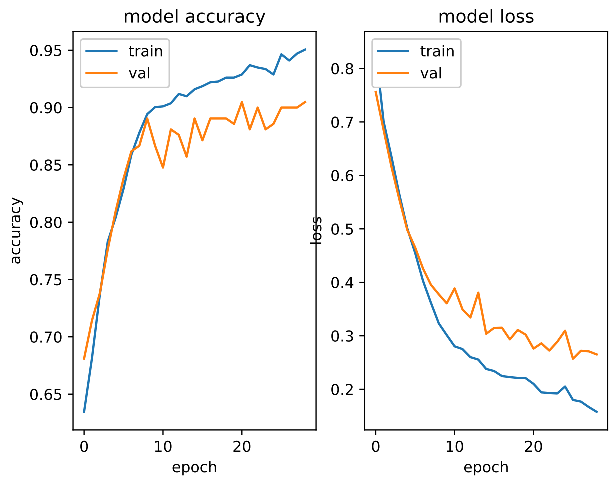

Fig. 5: Training Data Plots for the Base CNN Model for 1s Split

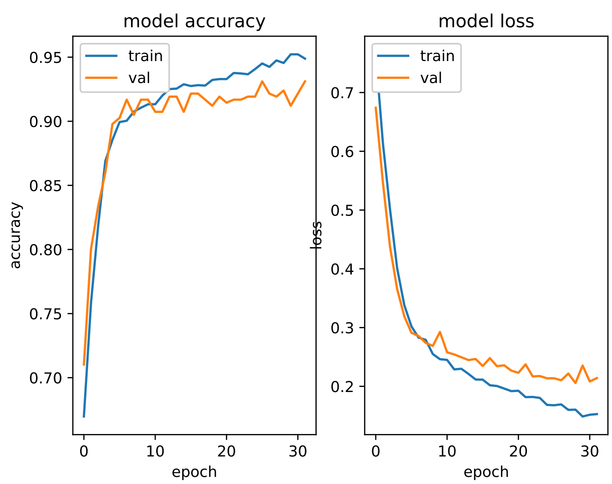

Fig. 6: Training Data Plots for the Base CNN Model for 0.5s Split

idation errors are better for 0.5s split. We summarize our results in the Table 3.

Table 3: Table of Results

Model and Split	Loss	Cat. Acc.	Precision	Recall
Base CNN with 1s split	0.2696	**0.9223**	**0.9280**	**0.9106**
Base CNN with 0.5s split	**0.2245**	0.9138	0.9186	0.9055
VGG16 with 1s split	0.5790	0.9082	0.9082	0.9082
VGG16 with 0.5s split	0.4309	0.8878	0.8895	0.8843

5 Conclusions

In this paper, we detail a novel way to classify chicken vocalizations into three classes - egg-laying, feeding, and distress. We show that by using a simplified CNN model and a split of 1s, we achieve an accuracy of 92.23%. Our method, since it does not delve into too much complexity, is easily implementable and deployable in any real-world scenario and could thus prove to be a step in the right direction when it comes to poultry health monitoring.

Considering we do not have a direct baseline to compare with, we can thus provide suggestions on how best to use our methods. We think that more advanced techniques of audio processing could lead to a cleaner audio, and thus would improve on the quality of the dataset. However, the biggest scope of research lies in better methodologies to collect the data of the chicken vocalisations without too much intrusion or influence. We had to collect data manually, which is a costly and time taking process - some sort of automation and standarized process would go a long way to solving this problem.

That being said, we believe that instead of aiming for accuracy, we must aim instead for generalizability, and thus more and more work using data from cheap recorders or noisy scenarios should be encouraged. Our method, being rooted in this philosophy, automatically becomes more practical to recreate, deploy and understand, while also attaining a good accuracy. We believe our work can help farmers who do not have the experience to understand chicken vocalisations provide better care to his farm, thereby leading to higher productivity rates and a fully automated poultry farm.

6 Acknowledgements

This work is supported by IIT Ropar Technology and Innovation Foundation (iHub – AWaDH) established by the Department of Science & Technology (DST), Government of India, at IIT Ropar in the framework of National Mission on Interdisciplinary Cyber-Physical Systems (NM - ICPS).

References

1. Brandon T. Carroll, David V. Anderson, Wayne Daley, Simeon D. Harbert, Douglas F. Britton, and Mark W. Jackwood. Detecting symptoms of diseases in poultry through audio signal processing. *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1132–1135, 2014.
2. Kaixuan Cuan, Tiemin Zhang, Junduan Huang, Cheng Fang, and Yun Guan. Detection of avian influenza-infected chickens based on a chicken sound convolutional neural network. *Computers and Electronics in Agriculture*, 178:105688, 2020.
3. Ryan R. Curtin, Wayne Daley, and David V. Anderson. Classifying broiler chicken condition using audio data. In *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1141–1144, 2014.
4. IMARC. India poultry market share, outlook and industry analysis 2023-2028.
5. Nikša Jakovljević, Nina Maljković, Dragiša Mišković, Petar Knežević, and Vlado Delić. A broiler stress detection system based on audio signal processing. In *2019 27th Telecommunications Forum (TELFOR)*, pages 1–4, 2019.
6. Axiu Mao, Claire SE Giraudet, Kai Liu, Inês De Almeida Nolasco, Zhiqin Xie, Zhixun Xie, Yue Gao, James Theobald, Devaki Bhatta, Rebecca Stewart, et al. Automated identification of chicken distress vocalizations using deep learning models. *Journal of the Royal Society Interface*, 19(191):20210921, 2022.
7. Kevin Mayer, Keith Ellis, and Ken Taylor. Cattle health monitoring using wireless sensor networks. In *Proceedings of the Communication and Computer Networks Conference (CCN 2004)*, pages 8–10. ACTA Press Calgary, AB, Canada, 2004.
8. Brian McFee, Colin Raffel, Dawen Liang, Daniel Ellis, Matt Mcvicar, Eric Battenberg, and Oriol Nieto. librosa: Audio and music signal analysis in python. pages 18–24, 01 2015.
9. Michael Mcloughlin, Rebecca Stewart, and Alan McElligott. Automated bioacoustics: Methods in ecology and conservation and their potential for animal welfare monitoring. *Journal of The Royal Society Interface*, 16, 06 2019.
10. A Arun Gnana Raj and J Gnana Jayanthi. Iot-based real-time poultry monitoring and health status identification. In *2018 11th International Symposium on Mechatronics and its Applications (ISMA)*, pages 1–7. IEEE, 2018.
11. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
12. Pieter Thomas, Tomasz Grzywalski, Yuanbo Hou, Patricia Soster De Carvalho, Maarten De Gussem, Gunther Antonissen, Frank Tuyttens, Eli De Poorter, and Paul Devos1 Dick Botteldooren. Using a neural network based vocalization detector for broiler welfare monitoring. 2023.
13. Huib van den Heuvel, Ali Youssef, Lisette M Grat, and Suresh Neethirajan. Quantifying the effect of an acute stressor in laying hens using thermographic imaging and vocalisations. *bioRxiv*, 2022.