BIODCASE TASK3: TINYML MODELS FOR BIRD CLASSIFICATION

Technical Report

Sebastián Espitia Londoño

Human-Environment Research (HER), La Salle Campus Barcelona — Universitat Ramon Llull, Spain

ABSTRACT

This technical report is given to participate in task 3 BioDCASE 2025. The task consists in identifying Yellowhammer birds, found throughout Europe. This task is tackled with two models, one trained with augmented data and another trained without any augmentation data, in this way, we can compare the behavior of both models. The model demonstrates high confidence in its predictions, with 95.96% of outputs classified as high-confidence decisions. Non-Enriched Model achieves superior feature learning while maintaining high confidence levels, suggesting better balance between certainty and complexity capture. Its enhanced clustering quality, combined with more calibrated uncertainty levels, positions it as the more robust choice for applications requiring both reliable predictions and meaningful internal representations. The trade-off in prediction consistency appears acceptable given the substantial improvements in representation quality.

***Index Terms*—** CONVOLUTIONAL NEURAL NETWORKS, MobileNet, Machine Learning, Augmented Data, Mel-Spectrograms

1. INTRODUCTION

This technical report is given to participate in task 3 BioDCASE 2025. This task consists in identifying Yellowhammer birds, found throughout Europe. The characteristics of the data set are given at [1]. It is important to note that the background of the data set changes between grasslands, forests. This is tackled with two models, one trained with augmented data and another trained without any augmentation data,in this way, we can compare the behavior of both models.

2. DATA PROCESSING

The pre-processing done is the baseline given for the project. For data augmentation, we have added two modifications to the original audio, a pseudo-random shift on tonality, from 2 semitones down to 2 semitone up, followed of a white noise addition [2]. This is recommended when training new models for a specific task [3] [4]. After these modifications, a Mel spectrogram of 40 channels is extracted from the loudest part of the audio, this to start classifying these audios from a spectral angle. Since the mel spectrogram is an image, the model used is a model that can classify images.

3. ENRICHED MODEL ARCHITECTURE

In this section, we will present the architecture of the model, which consists of a MobileNet model followed by two depth-wise convolutional networks and finally a with only two outputs, that will help classify the model outputs.

Table 1: MobileNet Slimmed V2 Enriched Model architecture Summary

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 55, 40, 1)	0
conv1 (Conv2D)	(None, 11, 20, 16)	640
conv1_bn (BatchNormalization)	(None, 11, 20, 16)	64
conv1_relu (ReLU)	(None, 11, 20, 16)	0
conv_dw_1 (DepthwiseConv2D)	(None, 11, 20, 16)	144
conv_dw_1_bn (BatchNormalization)	(None, 11, 20, 16)	64
conv_dw_1_relu (ReLU)	(None, 11, 20, 16)	0
conv_pw_1 (Conv2D)	(None, 11, 20, 16)	256
conv_pw_1_bn (BatchNormalization)	(None, 11, 20, 16)	64
conv_pw_1_relu (ReLU)	(None, 11, 20, 16)	0
spatial_dropout2d (SpatialDropout2D)	(None, 11, 20, 16)	0
conv_dw_2 (DepthwiseConv2D)	(None, 11, 20, 16)	144
conv_dw_2_bn (BatchNormalization)	(None, 11, 20, 16)	64
conv_dw_2_relu (ReLU)	(None, 11, 20, 16)	0
conv_pw_2 (Conv2D)	(None, 11, 20, 32)	512
conv_pw_2_bn (BatchNormalization)	(None, 11, 20, 32)	128
conv_pw_2_relu (ReLU)	(None, 11, 20, 32)	0
spatial_dropout2d_1 (SpatialDropout2D)	(None, 11, 20, 32)	0
conv_dw_3 (DepthwiseConv2D)	(None, 11, 20, 32)	288
conv_dw_3_bn (BatchNormalization)	(None, 11, 20, 32)	128
conv_dw_3_relu (ReLU)	(None, 11, 20, 32)	0
conv_pw_3 (Conv2D)	(None, 11, 20, 32)	1,024
conv_pw_3_bn (BatchNormalization)	(None, 11, 20, 32)	128
conv_pw_3_relu (ReLU)	(None, 11, 20, 32)	0
global_max_pooling2d (GlobalMaxPooling2D)	(None, 32)	0
dropout1 (Dropout)	(None, 32)	0
dense (Dense)	(None, 64)	2,112
batch_normalization (BatchNormalization)	(None, 64)	256
activation (Activation)	(None, 64)	0
dropout (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 2)	130
Total params:		**17,544 (68.54 KB)**
Trainable params:		**5,698 (22.26 KB)**
Non-trainable params:		**448 (1.75 KB)**
Optimizer params:		**11,398 (44.53 KB)**

The backbone of the model is taken from [5]. Since both models are equal in architecture but different by training, we will focus on the result.

4. RESULTS

4.1. Non-Enriched Model

Here we will showcase the results during training of the model with non-enriched data.

As the figure 1 indicates, an accuracy of 90.4%, the model has a high accuracy, without learning completely the dataset.

- High confidence predictions: 95.96
- Average prediction entropy: 0.050
- Feature clustering quality: 0.398

Especial Thanks to La Salle University Ramon Llull

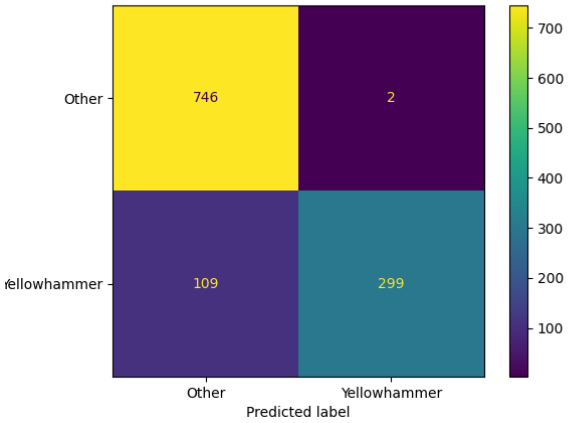

Figure 1: Confusion Matrix resulting for the non-enriched data augmentation.

- Prediction consistency (standard deviation): 0.134

The model demonstrates high confidence in its predictions, with 95.96% (2 a) of outputs classified as high-confidence decisions. The average prediction entropy of 0.050 (2 b), while low and approaching the theoretical minimum, indicates that the model maintains well-concentrated probability distributions with limited predictive uncertainty. This entropy level suggests the model exhibits appropriate confidence without extreme overconfidence.

The feature clustering quality score of 0.398 (2 c)indicates moderately successful learning of meaningful internal representations. This metric suggests that the learned features exhibit reasonable structural organization, with the model demonstrating capability in discovering coherent feature clusters within the data space. The clustering quality reflects a balanced approach to representation learning.

The prediction consistency, measured by a standard deviation of 0.134, demonstrates moderate variability in model outputs. While this indicates some fluctuation in predictions, the relatively contained variance suggests the model maintains reasonable stability across similar input conditions.

4.2. Enriched Model

Here we will showcase the results during training of the model with enriched data.

As the figure 3, the results are very promising as the accuracy is: 94% for the model. Suggesting that the model might have learned the dataset very well. even with an enriched one as we have made it. This can be due to the very nature of the dataset, which is explained here [6]. Now we will continue with the unsupervised results of the evaluation dataset:

The unsupervised evaluation of our model yielded the following quantitative metrics:

- High confidence predictions: 96.92%
- Average prediction entropy: 0.031
- Feature clustering quality: 0.375
- Prediction consistency (standard deviation): 0.115

A confidence prediction of 96.92% is consistent with the figure 4 a) which shows a great confidence of the model on the evaluation dataset. This finding is corroborated by the remarkably low average prediction entropy of 0.031, indicating that the model's probability distributions are highly concentrated rather than diffuse. The entropy value, approaching the theoretical minimum of 0, suggests the model exhibits minimal predictive uncertainty across the evaluation dataset.

The feature clustering quality score (4 c) of 0.375 indicates moderate success in learning meaningful internal representations. While this metric suggests that the learned features exhibit some degree of structural organization, the intermediate value indicates potential for improvement in the model's ability to discover coherent, well-separated feature clusters within the data space.

The prediction consistency, as measured by a standard deviation of 0.115, demonstrates relatively stable Non-Enriched Model behavior with moderate variability in outputs. This low variance suggests the model produces consistent predictions under similar input conditions.

5. COMPARISON OF MODELS

Table 2: Performance Metrics Comparison

Metric	Enriched Model	Non-Enriched Model	Advantage
High confidence predictions	98.24%	95.96%	Enriched Model
Average prediction entropy	0.031	0.050	Enriched Model
Feature clustering quality	0.375	0.398	Non-Enriched Model
Prediction consistency (std)	0.115	0.134	Enriched Model

Enriched Model achieves superior confidence metrics but may suffer from overconfidence, as evidenced by the extremely low entropy. Non-Enriched Model demonstrates more moderate confidence levels that may be better calibrated to actual prediction accuracy, suggesting more appropriate uncertainty quantification.

Non-Enriched Model significantly outperforms Enriched Model in feature clustering quality (0.398 vs 0.375), indicating superior ability to learn meaningful internal representations. This advantage suggests Non-Enriched Model better captures underlying data structures and complexities.

Enriched Model exhibits superior prediction consistency with lower variance, while Non-Enriched Model shows increased variability. This difference may reflect Enriched Model's tendency toward oversimplified decision boundaries versus Non-Enriched Model's more nuanced representational approach.

6. CONCLUSIONS

The comparative analysis reveals complementary strengths and distinct optimization strategies. Enriched Model excels in prediction confidence and consistency but may sacrifice representational richness. Non-Enriched Model achieves superior feature learning while maintaining high confidence levels, suggesting better balance between certainty and complexity capture. Non-Enriched Model's enhanced clustering quality, combined with more calibrated uncertainty levels, positions it as the more robust choice for applications requiring both reliable predictions and meaningful internal representations. The trade-off in prediction consistency appears acceptable given the substantial improvements in representation quality.

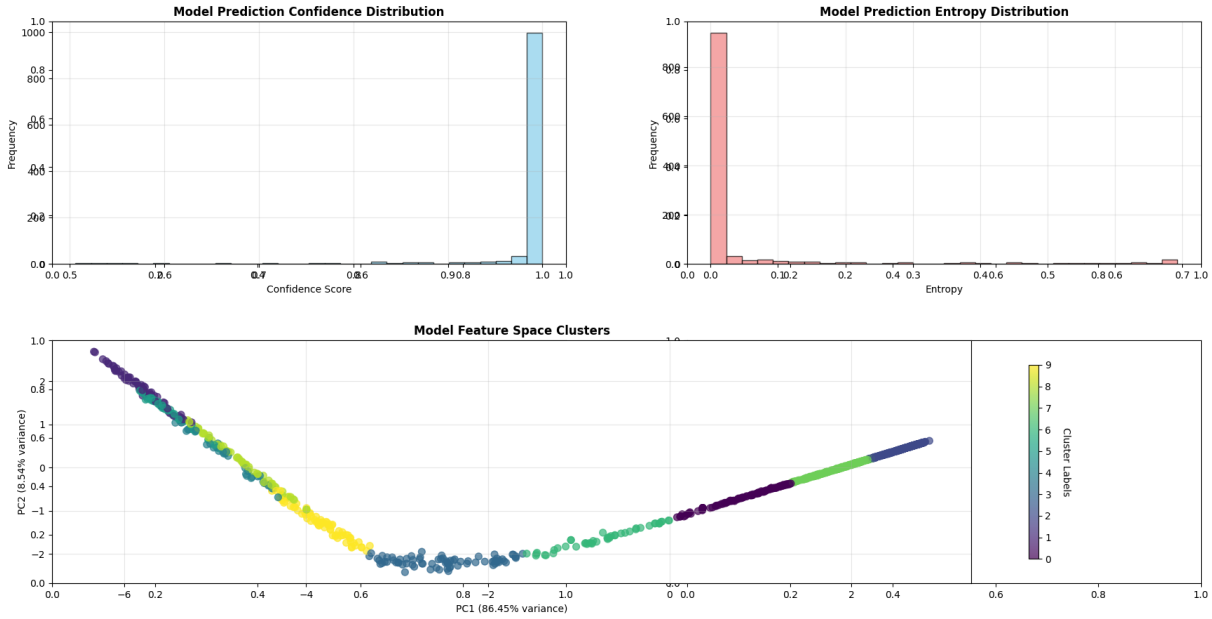

Figure 2: Results of the evaluation dataset: a) Confidence Distribution. b) Model Prediction Entropy Distribution. c) Model Space Clustering.

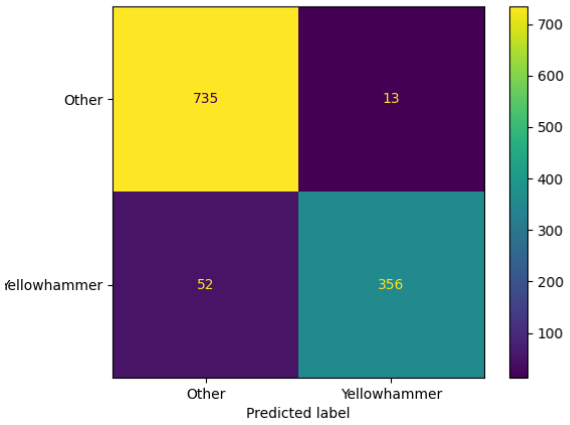

Figure 3: Confusion Matrix resulting for the enriched data augmentation.

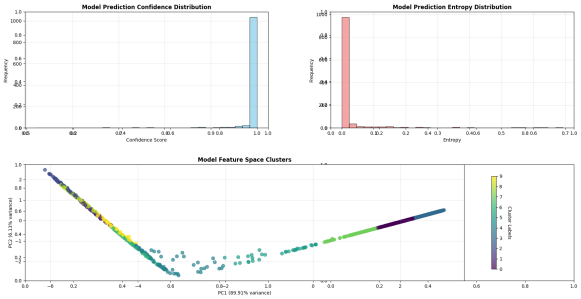

Figure 4: Results of the evaluation dataset: a) Confidence Distribution. b) Model Prediction Entropy Distribution. c) Model Space Clustering.

7. ACKNOWLEDGMENT

Special thanks for my wife, and my family for the support for this project. Also a mention to Ester Vidaña, whose knowledge was passed to me in order to create the models here mentioned.

8. REFERENCES

[1] https://biodcase.github.io/challenge2025/task3, 2025.

[2] https://kanerika.com/blogs/data-augmentation, 2025.

[3] A. J. N. Shermaine, M. Lazarou, and T. Stathaki, “Image compositing is all you need for data augmentation,” 2025. [Online]. Available: https://arxiv.org/abs/2502.13936

[4] S. Kumar, P. Asiamah, O. Jolaoso, and U. Esiowu, “Enhancing image classification with augmentation: Data augmentation techniques for improved image classification,” 2025. [Online]. Available: https://arxiv.org/abs/2502.18691

[5] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” 2017. [Online]. Available: https://arxiv.org/abs/1704.04861

[6] I. Morandi, P. Linhart, M. Kwak, and T. Petrusková, “Biodcase 2025 task 3: Bioacoustics for tiny hardware development set,” 2025. [Online]. Available: https://doi.org/10.5281/zenodo.15228365