CLASSIFICATION OF JAVANESE EAGLE TWEET BASED ON IMPROVED MEL-FREQUENCY CEPSTRAL COEFFICIENTS AND DEEP CONVOLUTIONAL NEURAL NETWORK

SILVESTER DIAN HANDY PERMANA

ASIA e UNIVERSITY

2024

CLASSIFICATION OF JAVANESE EAGLE TWEET BASED ON IMPROVED MEL-FREQUENCY CEPSTRAL COEFFICIENTS AND DEEP CONVOLUTIONAL NEURAL NETWORK

SILVESTER DIAN HANDY PERMANA

A Thesis Submitted to Asia e University in
Fulfilment of the Requirements for the
Degree of Doctor of Philosophy

April 2024

ABSTRACT

The Javanese Eagle is a rare and protected animal in Indonesia, threatened with extinction due to its limited population. Conservation efforts in zoos and nature reserves are essential to prevent their extinction. One critical aspect of conserving the Javanese Eagle is understanding their communication through tweets, which can provide insights into their needs and behaviours. This study addresses the problem of effectively classifying the Javanese Eagle's vocalizations to aid in their conservation. The primary technique involves the use of Improved Mel-Frequency Cepstral Coefficients (IMFCC) and Deep Convolutional Neural Networks (DCNN), combined to create a robust classification system. Data were collected from zoos and nature reserves in Indonesia, used to train and test the models, and then validated by experts. Experts validate after the best model is obtained and use new data to test its validity. The classification system aimed to distinguish between tweets indicating lack of food or drink, normal tweets, and those related to finding a partner. The study compared various CNN architectures, including AlexNet and VGGNet, and different combinations of training, validation, and test data. The best-performing model, VGGNet, was trained with a dataset split into 80% training, 10% validation, and 10% testing. During training, the VGGNet model achieved a peak accuracy of 100%, and during testing, it attained an accuracy of 99%. The Receiver Operating Characteristic (ROC) Curve analysis showed that the 'Normal' category had an area under the curve of 0.996, the 'Looking for Partner' category had an area under the curve of 1.000, and the 'Looking for Food' category had an area under the curve of 0.996. These results demonstrate the effectiveness of the proposed classification system in accurately identifying the Javanese Eagle's primary needs. The significance of this study lies in its potential to enhance conservation efforts by providing a reliable tool for monitoring the Javanese Eagle's well-being. By accurately classifying their vocalizations, conservation site managers can better understand and address the eagles' needs, improving their chances of survival and preventing extinction. This research also contributes to the broader field of bioacoustics and wildlife conservation, offering a methodology that can be adapted for other endangered species.

Keywords: Improved MFCC, deep convolutional neural network, Javanese eagle sound, sound classification

APPROVAL

This is to certify that this thesis conforms to acceptable standards of scholarly presentation and is fully adequate, in quality and scope, for the fulfilment of the requirements for the degree of Doctor of Philosophy.

The student has been supervised by: **Prof Ts Dr Titik Khawa Abdul Rahman**

The thesis has been examined and endorsed by:

Prof Ts Dr Ariffin bin Abdul Mutalib
Asia e University (AeU)
Examiner 1

Assoc Prof Ts Dr Suzaimah Ramli
Universiti Pertahanan Nasional Malaysia (UPNM)
Examiner 2

This thesis was submitted to Asia e University and is accepted as fulfilment of the requirements for the degree of Doctor of Philosophy.

……………………………….
Professor Dr. Siow Heng Loke
Asia e University
Chairman, Examination Committee
30 April 2024

DECLARATION

I hereby declare that the thesis submitted in fulfilment of the PhD degree is my own work and that all contributions from any other persons or sources are properly and duly cited. I further declare that the material has not been submitted either in whole or in part, for a degree at this or any other university. In making this declaration, I understand and acknowledge any breaches in this declaration constitute academic misconduct, which may result in my expulsion from the programme and/or exclusion from the award of the degree.

Name: Silvester Dian Handy Permana

Signature of Candidate: **Date**: 30 April 2024

ACKNOWLEDGEMENTS

Embarking on the journey of a PhD has been a transformative experience that would not have been possible without the support and encouragement of many remarkable individuals and institutions.

First and foremost, I express my deepest gratitude to God Jesus Christ for His divine guidance, blessings, and grace throughout this journey.

I would like to thank Prof. Dr. Titik Khawa Abdul Rahman who helped me and guided my thesis.

I am profoundly grateful to Universitas Trilogi. I thank the Rector, Prof. Dr Pramono Hariadi, and the Vice-Rector, Dr Anies Lastiati and Dr Dendi A. Gumilang. I also appreciate the former Rector, Prof. Mudrajad, and Vice-Rector, Prof. Indra Bastian.

To my beloved family: my wife, Seftya, and my child, Xavier, my parents, and my siblings, your support and belief in me have been invaluable.

Thanks to my research assistant, Mr. Fariz, and the dedicated team at Pusat Satwa Suaka Elang Jawa (PSSEJ).

I want to extend my appreciation to my PhD team from Universitas Trilogi: Mrs Rizka Ramayanti, Mr Budi Arifitama, Mr Yaddarabullah, Mrs Khoirina, Mr Rudi, Mr Umar Al Faruq, and Mrs Nina.

My deepest thanks to my dean, Dr Dina Nurul Fitria, and to my colleagues in the Informatics Department: Mr Yaddarabullah, Mr Budi Arifitama, Mr Ade Syahputra, Mr Ketut Bayu YB, and Mrs Dewi Lestari.

Lastly, I am deeply grateful to my team at the Human Resources Department: Mr Eko TS, Ms Ely Yulianti, and Mrs Tsani H. Saidah.

Thank you all for your contributions and for being part of this incredible journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

CNN	Convolutional Neural Network
MFCC	Mel Frequency Cepstral Coefficients
IMFCC	Improved Mel Frequency Cepstral Coefficients
DCNN	Deep Convolutional Neural Network
VGGNet	Very Deep Convolutional Networks
ROC	Receiver Operating Characteristic
ICT	Information and Communication Technology
MLPs	Multilayer Perceptron’s
FFT	Fast Fourier Transform
DFT	Discrete Fourier Transform
PSSEJ	Pusat Suaka Satwa Elang Jawa / Javan Eagle Wildlife Sanctuary Center
CQT	Constant-Q Transform
STFT	Short-Time Fourier Transform
SNR	Signal to Noise Ratio
ESC	Environmental Sound Classification
SSL	Self-Supervised Learning
HMM	Hidden Markov Models
TAR	True Acceptance Rate
TRR	True Rejection Rate
WAV	Waveform Audio Format
MFW	Mel-Frequency Wrapping
dB	Decibel

DPI	Dots Per Inch
RGB	Red Green Blue
ReLu	Rectified Linear Unit
LRN	Local Response Normalization
GPUs	Graphics Processing Units
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
AUC	Area Under the Curve
MAPE	Mean Absolute Percentage Error
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
Fmin	Minimum Frequency
Hz	Hertz

CHAPTER 1

INTRODUCTION

This chapter discuss on the Javanese Eagle whose existence is currently being threatened with extinction. In helping to preserve the Javanese Eagle, research is needed to identify the needs of the Javanese Eagle. Javanese eagles communicate using their tweets. From the tweet, it can be identified the conditions needs of the Javanese Eagle. The needs of the identified Javanese Eagle can help in the care of the Javanese Eagle. Research that can identify the needs of the Javanese Eagle from the sound of its tweets can help care taker provide the bird's needs in real-time.

The Javanese Eagle (Nisaetus bartelsi) is a rare and protected animal in Indonesia. These animals only live a few species globally and are threatened with extinction (Karpyn et al., 2020; Lindhout & Reniers, 2020). The Javanese Eagle is one of the animals that are conserved in zoos and nature reserves. These birds need to be bred to avoid extinction (Rose et al., 2017). Especially in zoos, caretakers need to pay attention to the needs of these birds, especially in maintaining a balance in nutrition. Balanced nutrition keeps the Javanese Eagle to survive. This Javanese Eagle is a rare and endangered species and is currently on the verge of extinction (Putra, 2015). They need to be taken very seriously to preserve their existence. One of the threats to the survival of the Javanese Eagle is its diminishing habitat. The Kamojang Eagle Conservation Manager is also an eagle observer said, all types of eagles in Indonesia are almost extinct. Even though in 1990, eagles were protected by the government, there are still many who trade eagles illegally (Putra, 2015).

Javanese eagles can communicate with one another by the sound of their tweets. The sound of the Javanese eagle tweet can indicate whether the bird is looking for prey, it is in normal condition, or even invites other Javanese eagles to breed. The voice of this tweet is very distinctive and very specific which can be heard (Kettler & Carr, 2019), (Berger et al., 2018).

The chirping sound of this Javanese Eagle can be studied and classified to help in the conservation of endangered animals. With the tweets studied by the proposed technique and verified by experts, can know the basic needs of the bird especially in searching for prey. This study will classify the sound of the Javanese Eagle for the benefit of animal conservation. The data from this study were taken from zoos and nature reserves in Indonesia and validated by experts. Data in the form of tweets will be classified. This research will develop a Javanese Eagle's sound classification technique that will classify the sound of the Javanese Eagle into lack of food or drink, knowing the Javanese Eagle in search of a partner, and normal state of bird tweet's through combination of algorithms from Mel-Frequency Cepstral Coefficients (MFCC) (Paul et al., 2021) and Deep Conventional Neutral Network (Murat et al., 2020; Niemi & Tanttu, 2018; Song & Li, 2019; Xie & Zhu, 2019). The results of this study can help managers of conservation sites in preserving the Javanese Eagle from extinction.

1.0 Background of the Study

Information and Communication Technology (ICT) offers various advanced techniques leveraged for diverse applications, including conservation efforts. This study applies ICT methods to classify and interpret the vocalizations of the Javanese eagle (Nisaetus Bartelsi). As a medium-sized eagle from the family Accipitridae, this species has seen its population decline due to habitat loss from volcanic eruptions and illegal hunting (Utami, 2021). Understanding and meeting the nutritional and environmental needs of these eagles are crucial for their conservation, particularly given their limited reproductive rate of one to two eggs per year.

Signal processing, particularly using Mel-Frequency Cepstral Coefficients (MFCC), plays a pivotal role in extracting distinctive features from sound signals (Abdul & Al-Talabani, 2022). MFCC transforms sound waves into parameters that encapsulate the characteristics of audio files, facilitating their analysis and classification. This research also improved the MFCC into IMFCC which further refines feature extraction by creating detailed feature vectors that capture the nuances of voice signals.

The developed Improved MFCC is combined with Convolutional Neural Networks (CNNs), a type of deep neural network which adept at the image processing two-dimensional data, such as images. CNNs are particularly effective in image recognition tasks due to their ability to learn complex features from data. However, since Multilayer Perceptrons (MLPs) are inadequate for handling spatial information in image data, the Javanese Eagle's sounds are first converted into spectrograms—a visual representation of the sound spectrum over time. Spectrograms provide a detailed picture of sound frequencies, enabling CNNs to train on this data more effectively and produce accurate classification models.

As an essential method in signal processing namely the Mel Frequency Cepstral Coefficient (MFCC) serves as a powerful tool for extracting distinctive features from sound signals. Through this technique, sound waves undergo transformation into various parameters, particularly cepstral coefficient parameters, which effectively encapsulate the characteristics of the audio file. This process plays a pivotal role in analysing and understanding the underlying properties of sound, facilitating tasks ranging from speech recognition to audio classification. Moreover, an advancement known as Improved MFCC further enhances the feature extraction process by generating comprehensive feature vectors from voice signals. These vectors encapsulate multiple dimensions of the audio, facilitating more nuanced analysis and recognition of speech features. Thus, the transformation of sound signals into spectrograms through the MFCC and its improved variants not only aids in understanding sound representations but also significantly contributes to the advancement of various applications in speech processing and beyond.

Image recognition techniques, especially those utilizing Convolutional Neural Networks (CNNs), excel in identifying patterns and features within images (Bharadiya, 2023). By converting audio signals into spectrograms, the same powerful CNN architectures used in image recognition were leveraged to classify the audio data. This approach bridges the gap between audio and image processing, enabling the use of advanced deep learning techniques which have proven successful in fields such as facial recognition, object detection, and scene analysis. Unlike Multilayer Perceptron (MLP), CNN is designed to process two-dimensional data, making it well-suited for image classification tasks. The high depth of the CNN network allows it to learn complex features from image data, making it a powerful tool for image recognition and classification.

A spectrogram is a visual representation of the spectrum of frequencies of a sound signal as they vary with time. It provides a specific picture of the sound image that CNN used to train adequately and produce an accurate model. The Mel Frequency Cepstral Coefficient (MFCC) model is a commonly used technique for converting audio data into a spectrogram. The MFCC model provided a specific picture of the sound image that CNN used for training adequately process in order to produce an accurate model.

Spectrograms provide higher accuracy in training than audio signals trained in digital form. By using spectrograms, CNN learned the complex features of sound data, making it a powerful tool for sound recognition and classification.

The Improved MFCC and followed by CNN in deep learning architecture was developed to classify the Javanese Eagle's in order to identify whether the Javanese eagles is lacking of food or drink, finding a partner, or it is a normal tweet of bird. The results of this research were used to help the bird’s caretakers to better understand the basic needs of the Javanese Eagle.

1.1 Problem Statement

The Mel-Frequency Cepstral Coefficients (MFCC) technique has been widely used in audio signal processing due to its effectiveness in feature extraction for speech and sound classification tasks. However, standard MFCC has notable limitations, particularly in recognizing high-frequency audio components (Kacur et al., 2022), which are crucial for accurately interpreting complex sounds such as the vocalizations of the Javanese Eagle. These limitations can lead to decreased accuracy in audio classification tasks, affecting the ability to effectively monitor and respond to the needs of these endangered birds.

In conservation efforts for the Javanese Eagle, recognizing specific vocal patterns is essential for timely and appropriate care. The inability of basic MFCC to accurately capture high-frequency audio details poses a significant challenge in this context. High-frequency sounds play a vital role in the communication of the Javanese Eagle, and their misinterpretation can lead to inadequate responses to the birds' nutritional and breeding needs.

To address the limitations of existing sound classification systems in recognizing the specific needs of the Javanese Eagle, this research conducted a thorough