SRI LANKAN ELEPHANT SOUND CLASSIFICATION USING DEEP LEARNING

Ariyasingha Gamage Hiruni Udarika Dewmini

219327R

Master of Science in Computer Science Specializing in Data Science, Analytics and Engineering

Department of Computer Science & Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

June 2024

SRI LANKAN ELEPHANT SOUND CLASSIFICATION USING DEEP LEARNING

Ariyasingha Gamage Hiruni Udarika Dewmini

219327R

Thesis submitted in partial fulfillment of the requirements for the degree
Master of Science in Computer Science Specializing in Data Science,
Analytics and Engineering

Department of Computer Science & Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

June 2024

DECLARATION

I declare that this is my own work and this Thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

The supervisor should certify the Thesis with the following declaration.

The above candidate has carried out research for the Master of Science in Computer Science Specializing in Data Science, Analytics and Engineering Thesis under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Prof. Dulani Meedeniya

Signature of the Supervisor: Date:

ACKNOWLEDGEMENT

I extend my sincere gratitude to Professor Dulani Meedeniya for her invaluable guidance and unwavering motivation throughout this research endeavor. Additionally, I express profound appreciation to my parents, friends, and the esteemed members of the Department of Computer Science and Engineering for their steadfast support and invaluable assistance, which have been instrumental in the realization of this work.

ABSTRACT

Understanding elephant caller types is crucial for various aspects of wildlife conservation and ecological research. By decoding the intricate vocalizations of elephants, researchers gain valuable insights into their behavior, social dynamics, and emotional expressions, which are pivotal for species conservation efforts. Elephant vocalizations serve as indicators of ecosystem health and vitality, aiding in ecological monitoring and biodiversity conservation initiatives. Furthermore, investigating caller types contributes to the preservation of cultural heritage by honoring the profound connection between humans and elephants across generations. In essence, delving into the world of elephant communication not only advances scientific knowledge but also fosters harmony between humans and these majestic animals, ensuring their long-term survival in the wild.

In this study, we delve into the domain of elephant caller-type classification utilizing raw audio format processing. Our focus lies on exploring lightweight models suitable for deployment on edge devices, including MobileNet, YAMNET, and RawNet, alongside introducing a novel model termed ElephantCallerNet, based on ACDnet architecture. Notably, our investigation reveals that the ACDnet-based ElephantCallerNet achieves an impressive accuracy of 89% when applied to a raw audio dataset. Leveraging Bayesian optimization techniques, we fine-tune crucial parameters such as learning rate, dropout, and kernel size, thereby enhancing model performance. Moreover, we scrutinize the efficacy of spectrogram-based training, a prevalent approach in animal sound classification. Through comparative analysis, we ascertain that for our dataset, raw audio processing outperforms spectrogram-based methods. In contrast to other models in the literature that primarily focus on a single caller type or binary classification (such as identifying whether a sound is an elephant voice or not), our models are designed to classify three distinct caller types: Roar, Rumble, and Trumpet. This approach significantly increases the complexity of our experiments compared to those discussed in the literature.

In the domain of elephant vocalization analysis, there has been limited exploration into the direct processing of raw audio data. Predominantly, various feature extraction techniques have been employed before training machine learning algorithms. In our investigation, we aim to bypass preprocessing stages and directly input raw audio data into machine learning models to assess the feasibility and efficacy of training on unprocessed audio signals.

Keywords: Elephant Vocalizations, Supervised-learning, Raw Audio Processing, Feature extraction, Classification

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
CNN	convolutional neural networks
DFT	Discrete Fourier Transform
ECN	ElephantCallerNet
FFT	Fast Fourier Transform
gPLP	generalized Perceptual Linear Prediction
HMM	Hidden Markov Model
HNR	Harmonics-to-Noise Ratio
LPC	Linear Predictive Coding
LSTM	Long Short-Term Memory
MCC	Matthews Correlation Coefficient
MFCCs	Mel-Frequency Cepstral Coefficients
SFEB	Spectral Feature Extraction Block
SSCs	Spectral Subband Centroids
STFT	short-time Fourier transform
SVM	Support Vector Machine
TFEB	Temporal Feature Extraction Block
YAMNet	*Yet Another Mobile Network*
ZCR	Zero-Crossing Rate

LIST OF APPENDICES