### Abstract

Excessive and intense barking has been a long-time problem for dog owners and households due to the harshness of the sound, the loud acoustic nature of the volume, and the fear of barking by a significant proportion of the population. The causes of dog barking are varied however barking caused by environmental noise has been particularly noticed. With the maturity of sound recognition technology, it has become possible to represent sound waveforms from Mel-frequency cepstral coefficients (MFCC) and Mel Spectrogram features to model and predict the time of appearance of barking dogs by a Convolutional Neural Network structured model. This paper will perform a classification model on possible dog bark triggers from public environmental sound databases and use the model to analyse and predict the barking behaviour from audio clips simulating a real-life dog barking scenario.

Keywords - Audio recognition, Convolutional Neural Networks, Mel Frequency Cepstral Coefficients, Dog Bark, Behaviour Prediction

### I. Introduction

**- Background information on the project**

In recent years, with the rapid development of deep learning, sound recognition has also been widely used in vary or fields, including automatic speech recognition (ASR)\cite{kabir2021survey} and environmental sound classification (ESC) \cite{bansal2022environmental}, and has been widely applied in smart home, medical, security monitoring fields. The emergence of audio recognition based on the development of machine learning and neural networks has made it possible to handle complex audio classification tasks. In addition to the areas mentioned, there has been quite a wide range of applications for animal voice recognition. The convolutional neural network (CNN) models outstanding ability to distinguish differences in audio of has made it possible perform breed identification and environmental quality issue based on animal sound analysis\cite{wang2019cough}.

Large numbers of automation of decisioning and classification of sound-based processing tasks have been raised and discussed. In this paper, a method to predict the cause of dog barking based on environmental sound recognition will be demonstrated. Due to the large variety and amount of noise in the city, these noises would cause anxiety in dogs and leads to excessive barking. Each individual dog is sensitive to different noises and lead to different reactions, as a result each dog will have their different noises that trigger barking depending on personality and environment. It is therefore natural for dog owners to seek to train their pets to minimize the behaviours that cause discomfort.

This prediction task of dog barking the following difficulties: (1) the nature of large amount and variety of urban noises around households therefore for general classification of noised a dataset contains large range of different noised is required; (2) difficulty to collect a large amount of data of each individual as the barking trigger sound is different, and there is almost no audio datasets about dog trigger barking is available, hence audio data augmentation technique is important; (3) based on breed, characteristic, growing environment etc. each dog would has different behavioural patterns, a general strategy to cover all to predict barking behaviour for a specific dog; (4) major limitation of the model about barking behaviour varies from each independent dog, and sound recognition could hardly cover all sources of barking triggers.

This paper discusses a machine learning-based approach to automate the prediction and barking behaviour of domesticated dogs to achieve behaviours that can intervene and prevent barking triggered by environmental noise and could further positively motivate the correct behaviour of dogs. Section 2 will go through a literature review of audio recognition, environmental sound classification systems, dog barking veterinary behaviour, and existing datasets; section 3 introduces the methodology used in the research, comparing different data processing, feature extraction, data augmentation techniques, CNN models and pre-trained model and hyperparameter tuning methods, managing to achieve a better performance in accuracy in efficiency; section 4 would evaluate the experiment result and demonstrate the model performance; section 5 discusses the limitation and potential future directions; and section 6 concludes the work and research. By conducting this research, an automated system for dispensing treats or snacks to the dog before it barks and using the model as a tool to encourage positive behaviour is aimed to be achieved.

### II. Literature Review

**-2.1 Techniques for sound classification**

The earliest research into voice recognition was scholars began to explore automatic speech recognition (ASR) in 1950s. Due to limitations in computing power as well as lack in sound analysis technology, the main methodology was pattern matching based on acoustic principles \cite{furui2010history}. The first big breakthrough of sound recognition was Hidden Markov models (HMMs) methods in the 1980s. The nature of excellent statistical power of HMMs models make it suitable and widely used in sound classification tasks. Later in the 1990s, the exponential increase in computer computing power and the widespread availability of audio files led to an explosion in automatic speech recognition including large vocabulary continuous speech recognition (LVCSR) systems. Now due to the advancements in deep learning in recent years, neural networks have demonstrated exceptional suitability for recognition tasks in effectiveness and accuracy, Support Vector Machines (SVMs) and convolutional neural networks (CNN) have become the predominant approach for voice recognition\cite{su2019environment}.

For sound feature extraction, common methods include MFCC, Gammatone, and LBP-HOG features.\cite{8096153}

**- 2.2 Overview of environmental sound classification**

Research on ambient sound has largely focused on the last 20 years or so when machine learning and neural networks have been widely used.

**- 2.3 Binary classification of sounds**

There is also extensive published work on the binary classification of sounds, with popular models including Support Vector Machine (SVM)\cite{muhammad2014pathological} and the CNN as mentioned above. In addition, numerous studies on biological characteristics classification have cited the effectiveness of random forest trees\cite{statnikov2008comprehensive} and logistic regression\cite{springer2015logistic} in similar classification tasks.

**- 2.4 Previous studies on dog behaviour and barking**

In existing research of dog behaviour, dog barking is not only a reaction to express anger, but also for many reasons. Studies have shown that dogs will bark when faced with fear, noise and separation anxiety, and female, neutered and more fearful dogs significantly responding to noise more strongly\cite{tiira2016prevalence}.

In addition, dog barking has been shown to have a role in communication. When dogs bark in communication and other behaviours, dog barking will be different from other dogs, and it will be affected by human preferences and social environments\cite{pongracz2010barking}.

**- 2.5 Existing datasets for environmental sound classification**

Environmental sound classification (ESC) is a vast research area, and numerous scholars have conducted extensive research in this field, and several related datasets are available. In recent researches the two databases widely used for environmental sound classification are UrbanSound8K\cite{} and ESC-50\cite{}. Two databases widely used for environmental sound processing are UrbanSound8K and ESC-50.

ESC-50 dataset

The UrbanSound8K dataset consists of a collection of urban sounds, which includes 8,732 common sounds heard in a city that also contains dog bark. Waveform signals will be extracted from all sound files for feature extraction. One disadvantage of this dataset is that the classification of environmental noise in real-life situations is not limited to those included in the dataset.

### III. Methodology

The high level approach would include a multi-class classifier will be developed using a CNN-based neural network model. The emphasis of this model will be on using appropriate convolution kernels to efficiently extract waveform signal features. To achieve the best training results for this classification task, various convolutional neural network structures, data inputs, number of pooling layers, and learning rate parameters will be explored.

The second classification task is a binary classification for audio, aimed at predicting whether a dog will bark. As each dog may behave differently, the training data will consist of combined sounds based on UrbanSound8k. Individual noises and barking sounds will be merged into longer audio files using Audio Clips. Several models will be built and compared for this classification task to select the best-performing one. The models to be tested include SVM, Random Forest, Logistic Regression, and CNN. Furthermore, techniques such as data augmentation, regularization, and hyperparameter tuning will be utilized to enhance the performance of the model. The incorporation of unsupervised learning methods may also prove effective for this classification task.

- Data collection and preprocessing

- Feature extraction methods

- Classification models (CNN, SVM, Random Forest, Logistic Regression)

- Model evaluation and comparison

- Data augmentation and regularization techniques

Since each audio file is lengthy, feature extraction becomes challenging due to the vast amount of data involved. To tackle this issue, our initial approach is to segment each audio file into smaller pieces and mark the class as ’N’ for the segments that the dogs didn’t react with. This technique will help reduce the size of each data piece while increasing the amount of data to facilitate the extraction of waveform feature graphics.

- Hyperparameter tuning

- Implementation details and software tools used

### IV. Results and Discussion

- Presentation of results for each classification task

- Evaluation of model performance

- Comparison of different models and techniques used

- Analysis of findings and discussion of insights

- Limitations and future research directions

**V. Conclusion and Recommendations**

**- Summary of the study**

Overall, this research signifies a direction in the use of machine learning techniques to understand and predict dog barking behaviour, which could be further refined and applied in real world scenarios such as in training aids or behaviour management tools for dogs.

**- Implications of the research for dog behaviour training**

Even though this study gives us some insight about dog barking triggers, it does have some shortcomings. The sound data we used only covers 10 types of urban sounds, which couldn’t cover sounds that dog might get attracted in a real life. Also, our measure of how good the model is at predicting barks is based on how well a sound classification model can label different sounds. If this model isn't very accurate, it could affect our model's ability to predict barking.

**- Recommendations for future research**

For future work, one direction is to involve real-time Acoustic Event Detection and unsupervised learning, that it could be able to train or manage a dog's behaviour in real time, changing based on the sounds around the dog. Another is to combine veterinary behaviour studies to improve our model. If we know more about the tendencies of specific breeds, or health status of individual dogs, we could make a completer and more accurate model for predicting when a dog is going to bark.

- Conclusion and final remarks