Beyond the Spectrum: Custom MFCC Processing For Acoustic Health Monitoring

Zhuolin Liu

A thesis submitted in partial fulfillment for the
degree of Master of Science
in
Electrical and Computer Engineering

APPROVED:

Professor Bashima Islam

Professor Edward A. Clancy

Professor Xinming Huang

Worcester Polytechnic Institute

Abstract

Mel-Frequency Cepstral Coefficients (MFCCs) are a critical feature in audio signal processing and have wide applications in systems that require audio classification, including the field of health monitoring and human activity recognition. The importance of MFCCs lies in their ability to mimic the human auditory system's response, making them particularly useful for analyzing audio signals in ways that are meaningful for classification tasks. The paper introduces a method for enhancing audio classification by developing custom Mel-Frequency Cepstral Coefficients (MFCC) for health monitoring and classification of human activities. A diverse audio dataset, focusing on a balanced subset to examine their periodograms, which involves calculating the magnitude squared of the frequency response. Identifying the predominant frequency in each audio file—defined as the frequency with the highest power—leads to the creation of an ordered array of these frequencies. This array, devoid of redundancies, is utilized to establish frequency bins for the MFCC algorithm, laying the groundwork for a customized filter bank automatically tailored to the dataset's specific characteristics.

The limitation of current MFCCs in handling extremely high or low frequencies becomes evident. This limitation is due to the logarithmic scale distribution of frequency bins, which results in a denser concentration of bins at lower frequencies, starting from 20Hz, and becoming progressively sparser towards the higher frequency limit of 20kHz. This characteristic of MFCCs underscores the need for careful consideration when employing this tool in sound analysis, particularly for health monitoring purposes.

To address the issue, a custom MFCCs approach is proposed. This customized approach significantly enhances audio classification system performance, as proven through extensive testing on various human-related audio datasets. These datasets include gender identification, environmental sounds, health-related sounds (like breath and vocal sounds for disease detection), and emotional speech analysis. The comparison between the traditional MFCC and the custom version shows a notable increase in classification accuracy, particularly with pitch-shifted audio samples. In order to test the results of Custom MFCC, we divided the dataset into train and test dataset in a form of 80% of train and 20% test randomly. Across 4 dataset, the average improvement of custom MFCC is 7.2% and with pitch-shifting is 6.7%. This indicates the custom MFCC's superior ability to handle human sound variations, highlighting its potential to improve audio classification tasks and its application in complex audio scenarios. Such advancements benefit a range of technologies that rely on sound analysis, marking a significant step forward in the field.

Acknowledgements

I would like to thank BASH LAB for their help, guidance, and assistance throughout the entirety of this project. Additionally we would like to thank Worcester Polytechnic Institute for providing us with the opportunity to complete this project. This thesis used human dataset that is publicly published online.

Contents

List of Figures

1 Introduction

Breath monitoring utilizes microphones to capture the distinct sound signatures produced by different breathing patterns. These patterns can indicate various respiratory conditions such as asthma, chronic obstructive pulmonary disease (COPD), or sleep apnea. Sophisticated signal processing algorithms like MFCC analyze these sounds to extract crucial features that help in diagnosing and monitoring respiratory issues. In heart rate monitoring, sensitive microphones equipped in wearable devices pick up the subtle sounds of the heart's mechanical movements, such as the valves opening and closing. Noise-reduction technologies are essential to isolate these heart sounds effectively. Once captured, MFCCs transform these sounds into a format where critical components are emphasized, allowing algorithms to analyze heart rate and variability.

Incorporating the analysis of both inhalation and exhalation sounds from adults and the crying sounds of babies into health assessments can provide invaluable insights into their respiratory health and emotional well-being. This method presents a non-invasive way to monitor and evaluate health indicators through sound, making it especially crucial for individuals lacking access to medical resources in remote or underserved areas. Moreover, sound signal analysis can extend to the detection of mental health conditions, where variations in vocal patterns, speech rate, and tone can indicate psychological states or stress levels. By leveraging sound analysis, early detection and intervention strategies for mental health conditions become more feasible, facilitating a broader approach to healthcare.

In addition to these applications, a critical application of microphones lies in audio processing for health monitoring, an area that has become increasingly relevant in emerging technologies. Specifically, concentrating on human sound classification and identification can significantly enhance capabilities in several key areas, including emotion recognition from vocal patterns, distinguishing between different speakers for security purposes, and refining

user interactions with voice-controlled devices. For health monitoring, one study discovered that MFCCs are the most commonly used feature[25]. However, while traditional Mel-Frequency Cepstral Coefficients (MFCC) work well for conventional applications, they may not be ideally suited for the unique environmental challenges of health applications. In health monitoring, the audio signals of interest, such as those generated by heartbeats or breathing, often occur in noisier and more complex acoustic environments. These environments demand more tailored audio processing techniques that can accurately isolate and analyze medically relevant sounds from background noise. Therefore, there is a growing need to develop custom MFCCs and other audio processing methods specifically designed for the nuanced demands of health monitoring to enhance accuracy and reliability in these applications.

Mel-Frequency Cepstral Coefficients (MFCCs) is one of the mostly common acoustic features used to analyze audio data, which are instrumental in analyzing the frequencies within the human audible spectrum, ranging from 20Hz to 20kHz. However, it's important to recognize the limitations of MFCCs in handling extremely high or low frequencies for health applications, since the breathing sound and heart beat sound are extremely low frequencies. This limitation is due to the logarithmic scale distribution of frequency bins, which results in a denser concentration of bins at lower frequencies, starting from 20Hz, and becoming progressively sparser towards the higher frequency limit of 20kHz. This characteristic of MFCCs underscores the need for careful consideration when employing this tool in sound analysis, particularly for health monitoring purposes.

To address this limitation, we introduces an innovative technique that transcends the traditional 20Hz to 20kHz frequency constraints by implementing a custom configuration of the Mel filter bank. This tailored approach aims to extend the effective range and resolution of MFCC analysis, ensuring a more equitable representation across the entire audible spectrum. By strategically positioning the Mel filter bank, our method enhances the capture of essential audio signal information, particularly in regions that are critical for the intended application but were previously marginalized by conventional MFCC processing.

1.1 Innovative Techniques in MFCC Processing

This method in MFCC processing becomes particularly significant when analyzing diverse human voice and activity datasets. The adaptation of our method showcases the adaptability and potential of this feature extraction technique in various applications such as, health monitoring and emotional detection. Specifically, the datasets such as the Respiratory Sound Database, RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song), and the DonateACry dataset.

We start by selecting a balanced subset from a dataset containing various classes of audio files. The periodogram for each file is calculated by determining the magnitude squared of the frequency response. We identify the predominant frequency in each file—the frequency with the highest power. These frequencies are then compiled into an ordered array, carefully refined to remove redundancies. This array forms the basis of the frequency bins, serving as the central frequencies in custom MFCC filter bank. This structured process efficiently transforms raw audio into a format crucial for custom MFCC computation, providing a systematic approach to audio data analysis. We utilize 4 distinct dataset, Respiratory Sound Database, RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song), and the DonateACry dataset, to evaluate the performance of our proposed custom MFCC. The results demonstrate that, across four distinct datasets, the accuracy of the test utilizing our custom MFCC algorithm showed a significant improvement of 13% when compared to the conventional MFCC approach.

The literature review section examines existing studies on voice disease detection and human activity classification, setting the stage for our contribution. Within the methodology section, we detail the Python packages utilized, and provide an in-depth exploration of the Mel Frequency Cepstral Coefficients (MFCC) process, explaining the conventional methodology for calculating MFCCs. Additionally, we analyzed alternative methods that fell short of expectations, discussing both their underlying rationale and their limitations. Subsequently, we introduce our custom MFCC approach, detailing both the theoretical foundation behind this method and the mathematical equations employed for frequency calculation. The custom MFCC function was detailed, including an explanation of its operation and the incorporation of pitch shifting to shift the acoustical signal frequency of the original dataset. Subsequently, a machine learning model was deployed to assess the performance of the custom MFCC method.

2 Literature Review and Background Study

This section describes the existing works on different meethod used for human activity, disease detection and sound detection.

2.1 Theoretical Background

The Mel-Frequency Cepstral Coefficients (MFCCs) are a feature that widely used in human speech and audio processing. The theoretical background of MFCCs is rooted in the principles of human sound perception and signal processing.

The design of MFCCs is deeply inspired by the intricacies of the human auditory system from 2 aspects. The first aspects, known as the Critical Band Theory, it shows that the human ear's ability to resolve frequencies varies along the frequency spectrum, which divides the spectrum into numerous critical bands. These bands are observed to be narrower at lower frequencies and wider at higher frequencies. The second aspect is the Mel Scale, which posits that human perception of frequency does not adhere to a linear scale, but rather follows a logarithmic one. This scale is instrumental in quantifying how pitches are perceived by listeners, specifically how pitches are judged to be equidistant from one another despite the actual logarithmic frequency spacing.

2.2 Understanding conventional MFCC

A Voice Disease Detection Method based on conventional MFCCs and Shallow CNN. This paper focuses on using software for remote diagnosis of voice diseases. It employs conventional MFCC parameters and a convolutional neural network (CNN) for analyzing voice samples from 352 patients, achieving up to 92% accuracy. This approach shows significant improvements in detecting voice diseases with increased accuracy and computa-

tional efficiency[16].

Voice in Parkinson's Disease: A Machine Learning Study. The paper focues on machine learning algorithms to analyze voice changes in Parkinson's disease (PD) patients at different disease stages and under different therapy conditions[17]. The extraction method they used called **Opensmile.**, which supports an extensive array of audio low-level descriptors like CHROMA, CENS features, loudness, conventional MFCCs, and various others, alongside the application of delta regression and statistical functionals to these descriptors for enhanced analysis[18].

Voice disorder classification using convolutional neural network based on deep transfer learning. Such paper focuses on real-time embedded system designed to enhance base calling in third-generation sequencing technologies. The paper used Mel spectrogram as input of OpenL3. The process involves resampling audio signals to 48 kHz, segmenting them into frames with a Hamming window, and transforming these into the frequency domain using the short-time Fourier transform (STFT). These frequency domain frames are then filtered through a Mel filter bank—composed of triangular filters that smooth the spectrum and condense data volume—to calculate Mel bands[22].

Voice Disorder Identification by Using Machine Learning Techniques. This research focuses on enhancing m-health systems for diagnosing and monitoring voice disorders, particularly dysphonia, through machine learning techniques. It cooperates the use of conventional MFCC among other acoustic parameters for distinguishing pathological voices from healthy ones[20].

Infant Cry Signal Diagnostic System Using Deep Learning and Fused Features. The paper introduce an approach to early disease detection using the only communication method available to infants: crying. Leveraging machine learning (ML) models for analyzing cry signals presents a noninvasive, efficient alternative to traditional diagnostic methods. This research aims to develop an automated diagnostic model that utilizes feature extraction

from cry signals to identify diseases like sepsis and RDS at their onset[21].

Artificial intelligence framework for heart disease classification from audio signals. This paper presents an approach for detecting heart disease using audio signals to optimize the detection technique. The approach invovles several steps: data acquisition, augmentation, pre-processing, feature extraction (including conventional MFCCs) among eight methods), feature normalization, model selection, implementation, and result prediction[24].

Lung disease recognition methods using audio-based analysis with machine learning. The paper under review delves into the advancements in lung sound-based diagnostics facilitated by computer-based automated approaches and enhancements in lung sound recording techniques. The method implemented using conventional MFCC, spectrogram properties, Short-time Fourier Transform (STFT), Mel-STFT, Empirical Mode Decomposition (EMD), and data de-noising. These extracted features are then analyzed using deep learning algorithms, which are particularly adept at handling unstructured data, allowing for the direct use of raw audio signals or their spectrograms. From their research, it was found that among the papers examined, 9.4% utilized the Short-Time Fourier Transform (STFT) approach for feature extraction. Other prevalent techniques included Mel spectrograms at 16.4%, log-mel spectrograms at 18.5%, and (conventional MFCCs) at 25.5%. The remaining 30% of the studies investigated various other methods, such as Zero-Crossing Rate (ZCR) energy, Constant Q Transform (CQT), and a bag of words approach, as referenced in Figure 2. This comprehensive review underscored the significance of feature extraction methods in sound classification, particularly in diagnosing lung and respiratory diseases. Specifically, the analysis revealed that STFT was employed in 9.4% of the analyzed papers, followed by Mel spectrograms, log-mel spectrograms, and conventional MFCCs, with the aforementioned percentages. The rest of the studies explored alternative techniques, including ZCR energy, CQT, and a bag of words, as detailed in Figure 1[25].

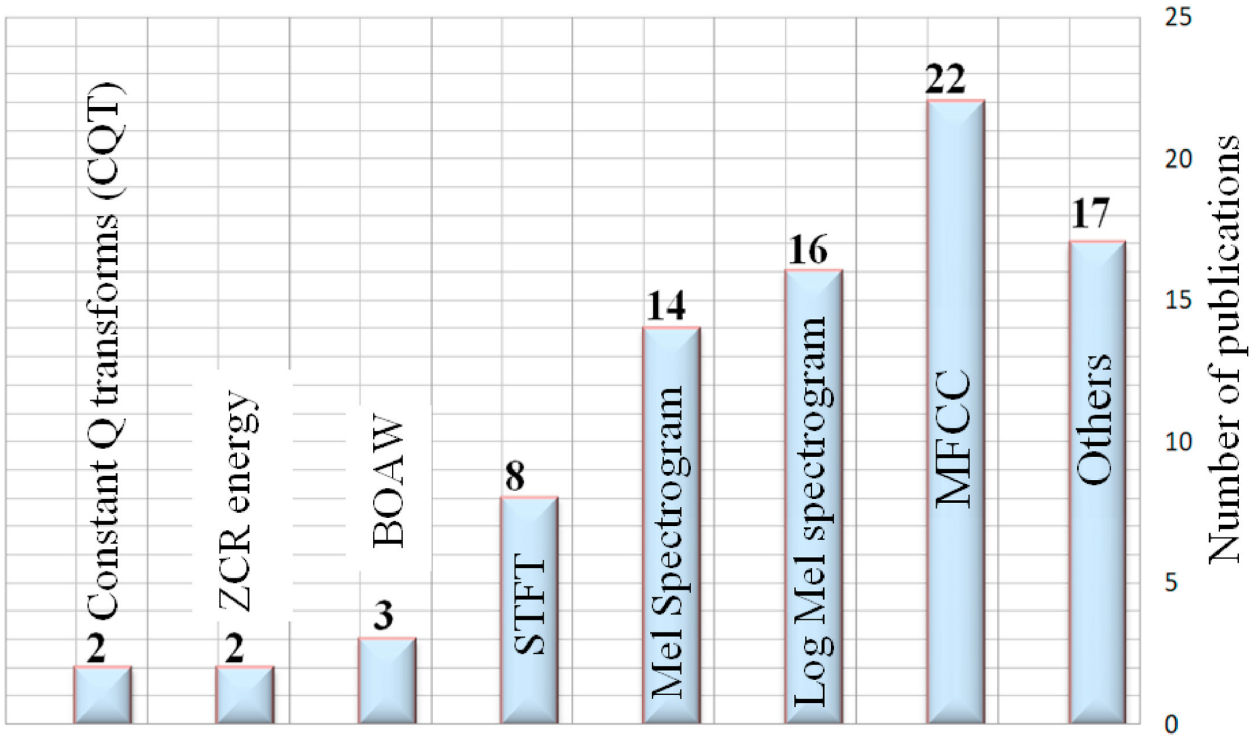


Figure 1: Statistics of feature extraction techniques in sound-based classification[25]

2.3 Human activity classification

Human activity classification based on sound recognition and residual convolutional neural network. This study highlights the significance of utilizing data on human activities to understand interactions between people and their environments, emphasizing the challenges of traditional data collection methods due to high costs and labor-intensive processes. it shows limitations of IMU and Computer Vision method, in environments with physical obstacles, insufficient lighting, or where wearing sensors is impractical. The study proposes a sound recognition-based model for classifying human activities, detailing the research methodology encompassing literature review, dataset construction, model development, and performance analysis[19].

Lightweight Audio-Based Human Activity Classification Using Transfer Learning. The paper outlines an approach to Human Activity Recognition (HAR) using audio

signals, with a focus on the application of audio-based recognition models for real-time classification on smartphones. They emploed YAMNet, a neural network architecture for audio classification, which processes log-Mel spectrograms to classify audio into one of 521 classes from the AudioSet-Youtube corpus. Based on MobileNet-v1, YAMNet is optimized for computational efficiency with 27 convolutional layers and a fully connected layer, of which only 28 layers have learnable weights[23].

3 Methodology

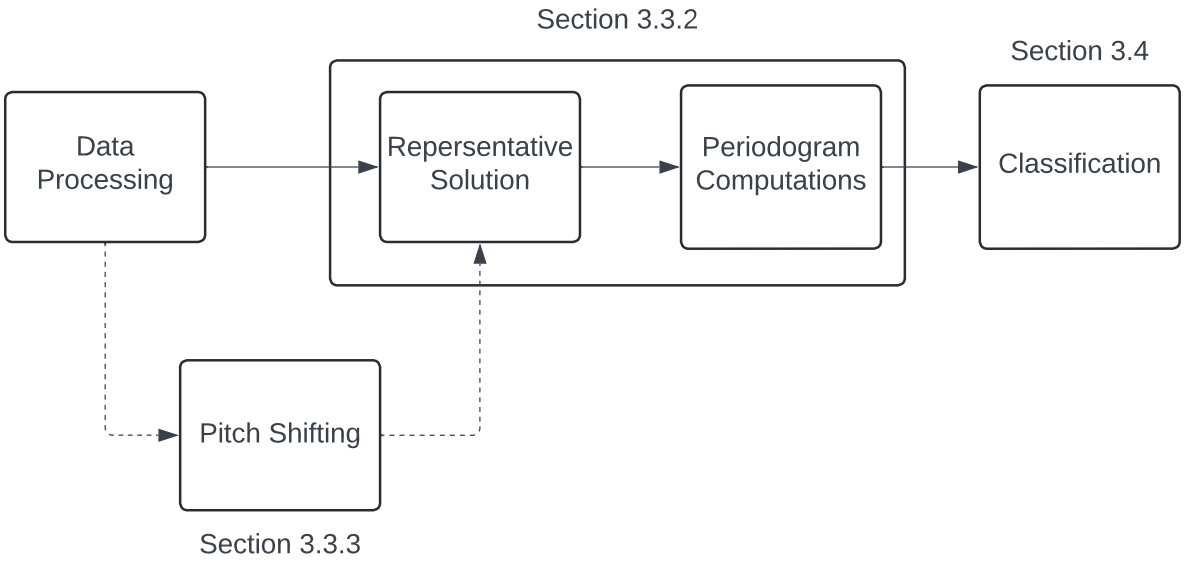


Figure 2: Block Diagram of Custom MFCC Calculation

This flowchart outlines the overall workflow of our proposed system. In section 3.3.2, we detail the customized computation of Mel Frequency Cepstral Coefficients (MFCC), starting from audio file preprocessing to the application within a machine learning model, as depicted in Figure 2. Our approach emphasizes representative selection of center frequencies, computation of the periodogram for each audio file, and identification of predominant frequencies essential for our custom MFCC function. In section 3.3.3, we introduced the con-

cept of pitch shifting, a technique that alters the pitch of audio data to higher frequencies. This method is instrumental in evaluating the performance differences between the conventional MFCC and our custom MFCC, allowing us to validate the enhanced capability of our system to handle broader frequency spectra. In section 3.4, we discussed the integration of CNN and LSTM networks is explored, demonstrating the model's effectiveness in capturing complex patterns and dependencies within audio signals. This hybrid model underscores the advanced capabilities of our system, setting a new standard for audio signal processing in research and practical applications.

3.1 General Approach

We begin by selecting a dataset consisting of audio files in different classes. We take an evenly distributed subset of these classes and find their periodogram. This is done by finding the magnitude squared of the frequency response. For each audio file, the predominant frequency is found by locating the frequency corresponding to the highest power present in the signal. Subsequently, these predominant frequencies are consolidated into an array. The formation of this array is carried out with meticulous attention, ensuring it is structured in ascending order and meticulously sorted to excise redundant information. This refined array metamorphoses into the list of frequency bins, which are essentially the center frequencies in the filter bank, pivotal for the MFCC algorithm. The overarching process for computing MFCC unfolds as outlined, embodying a methodical approach to transforming raw audio data into a structured and insightful representation.

- Window the data with a hamming window,
- Shift it into FFT order,
- Convert the FFT data into filter bank outputs,
- Find the log base 10,

- Find the cosine transform to reduce dimensionality [12].

The default MFCC filter bank involves placing triangular filters spaced out based on the Mel scale (logarithmic scale). Instead of following the default Mel scale, we implement a customized center frequencies found from the predominant frequency algorithm. The default MFCC filter bank is designed to perform well within the human hearing range, from 20Hz to 20kHz. The custom filter bank allows us to perform better within and beyond this range, making it more robust.

3.2 Alternative Methods Explored

In the quest to solve the limitation of current MFCC, A alternative methods were explored to identify the most effective solution.

3.2.1 Introduction to Alternative Method

Using K-means Clustering method to determine the dominant frequencies of the selected dataset. We extract the spectrogram by calculating Short time Fast Fourier Transform of each audio files within in the dataset to have a good representation of the power of the signal in each category. The goal of K-means Clustering is to find a way to partition n observations, in this case the spectrogram, into K clusters, in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells.

3.2.2 Rationale Behind Alternative Method

Audio datasets consist of diverse categories, each characterized by unique features and distinctions. After we calculate the spectrogram of each category, By analyzing the spectrogram, we can gain insights into which category exhibits the most dominant frequency

components. As we input the spectrogram into K mean clustering with number categories in the dataset as number of K, it supposes to give us good clustered frequencies as the output into Voronoi cells (see Figure 3).

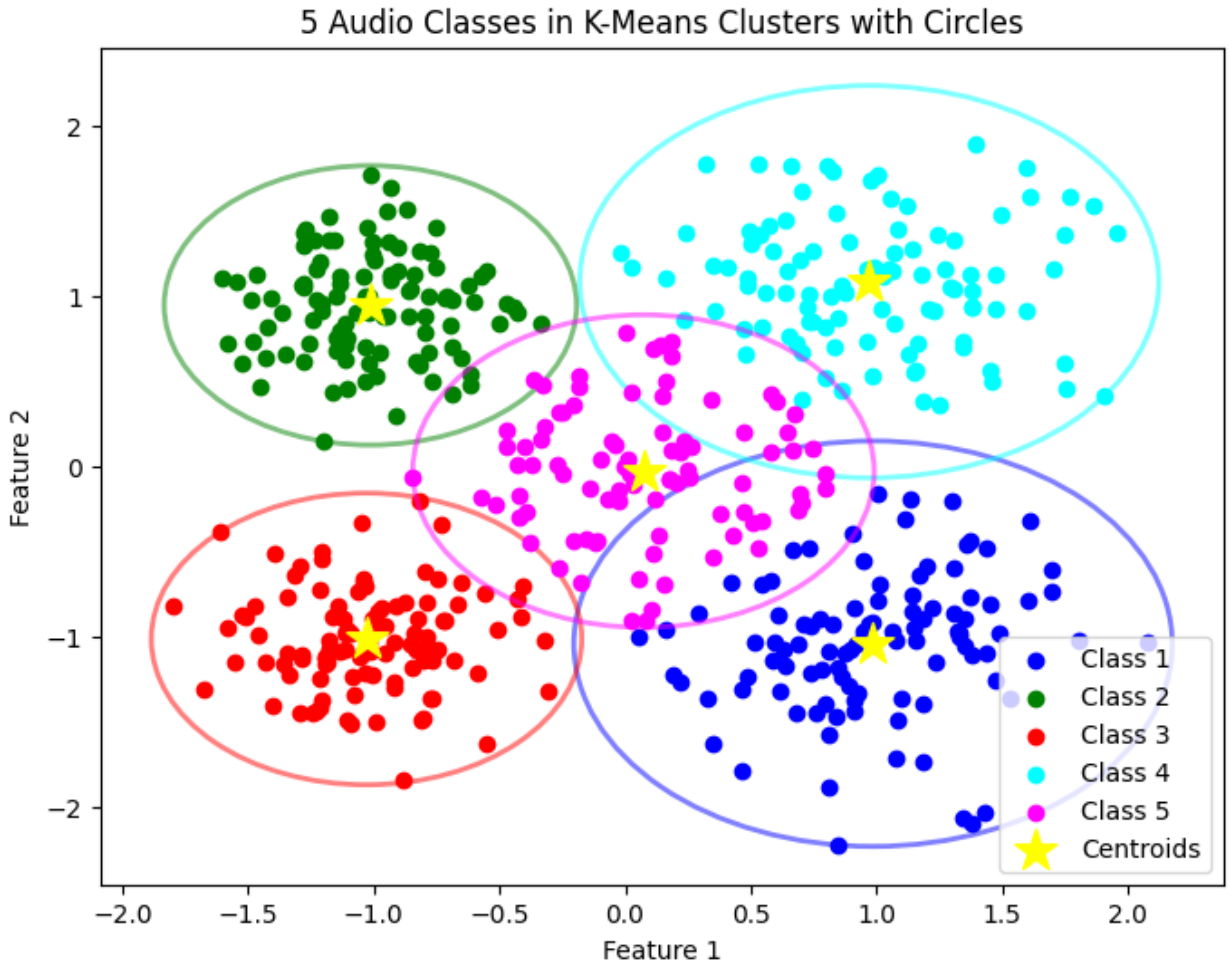


Figure 3: Ideal Situation In K-means Clustering

Once we implemented those ideal frequencies output from K-means Clustering. Next, we input those frequencies into a customized filter bank, the ideal triangle filter bank should look like Figure 4

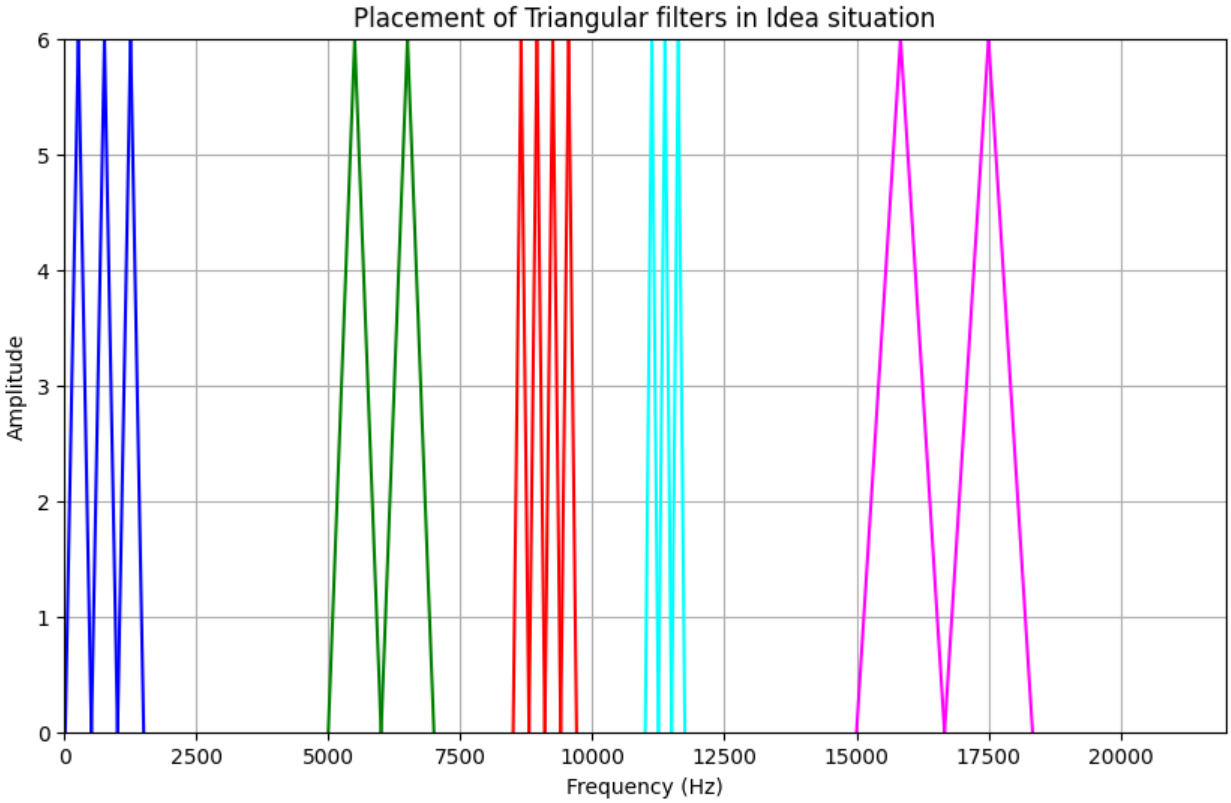


Figure 4: Ideal Situation In K-means Clustering of Triangular Filter Placement

3.2.3 Issue with Alternative Method

The limitation with the alternative approach lies in its divergence from real-world data. In real-world signals, significant overlap occurs between the frequencies of different categories within the dataset. For example, noise in each category may share the same frequency and carry significant weight. This overlap can lead K-means clustering algorithms to mistakenly perceive these shared frequencies as centroids for each dataset, resulting in misclassification of categories.

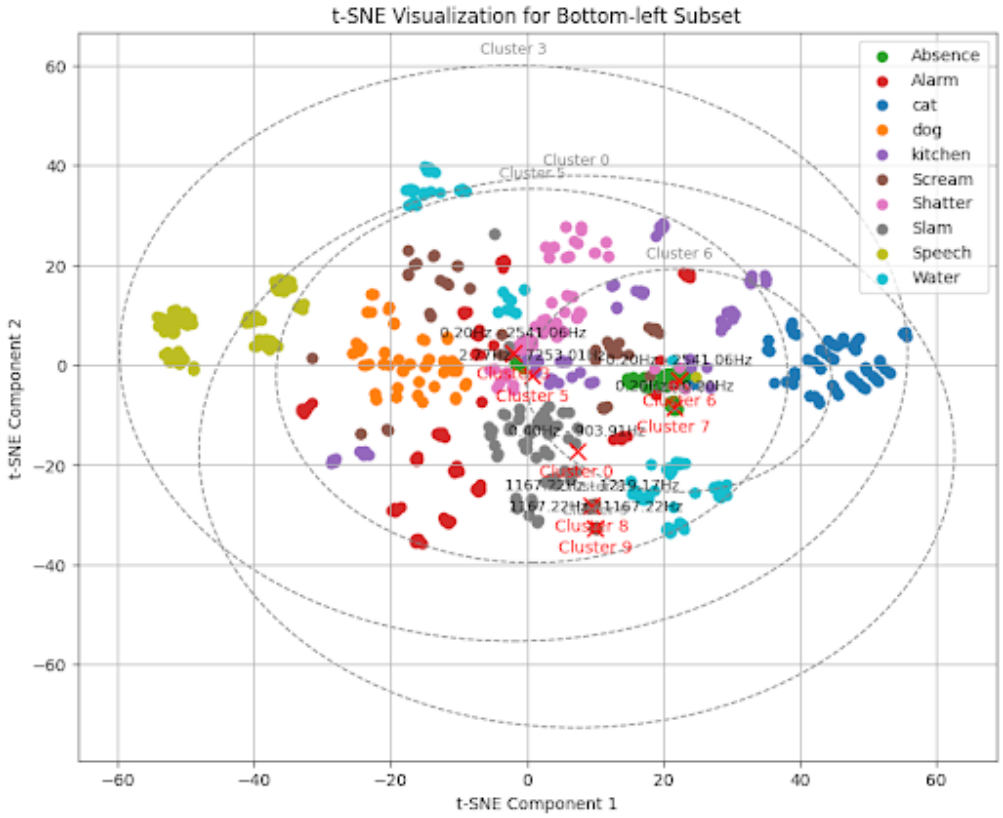
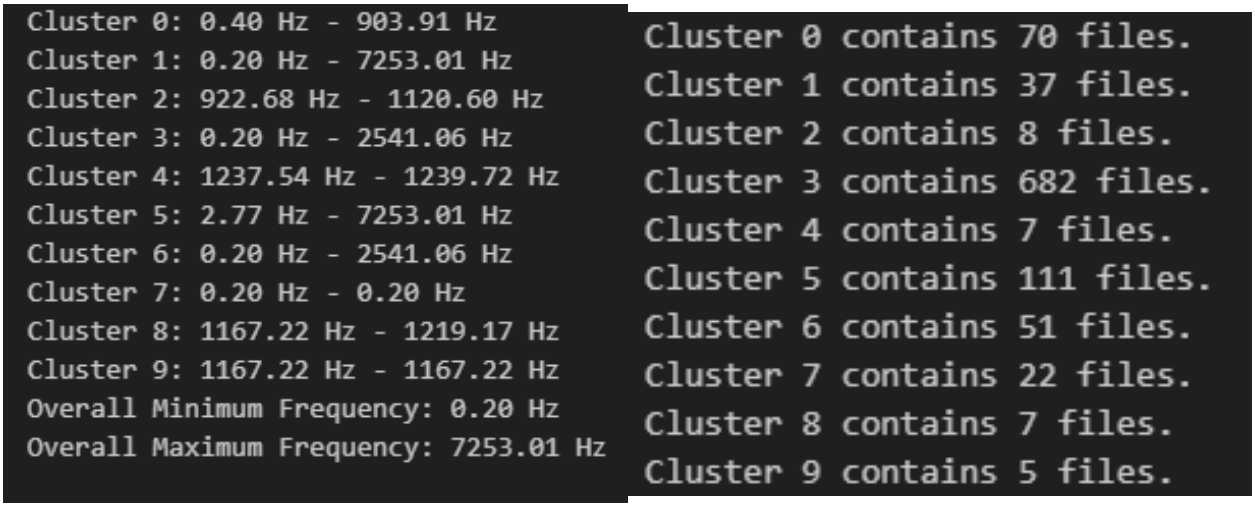


Figure 5: Empirical Data In K-means Clustering

In Figure 5, the clusters exhibit significant overlap, potentially causing the K-means Clustering algorithm to aggregate the majority of audio signal files into a single cluster. Figures **??** and **??** highlight this tendency, revealing that the bulk of audio files are assigned to cluster 3, which consequently holds the majority of the dataset. Moreover, after scaling, all triangular filters are constrained within the narrow frequency range of 0.2Hz to 2541Hz, effectively disregarding other frequencies



Frequency ranges

Number of Files in each Cluster

Figure 6: Empirical Data in K-means Clustering

As a result, the K-means clustering method may not accurately capture the nuanced distinctions present in real-world data. The tendency for significant overlap among frequency categories, as illustrated in Figure 5, challenges the algorithm's ability to discern distinct clusters. This limitation is further emphasized by Figures 6, which reveal a disproportionate allocation of audio files into cluster 3, indicative of potential misclassification issues.

Moreover, the confinement of triangular filters to a narrow frequency range, as observed post-scaling, shown in Figure 6, disregards other relevant frequencies. These constraints limit the algorithm's ability to accurately characterize the diverse frequency spectrum present in real-world audio signals. Therefore, while the K-means clustering method offers computational efficiency and simplicity, its performance may be compromised in scenarios where data exhibits complex overlapping patterns, as often encountered in audio signal processing tasks.

3.3 Proposed Methodology

Figure 2 presents the comprehensive workflow of the proposed method, detailing the sequence from preprocessing WAV files through to the application of a custom MFCC function within a machine learning model. This diagram illustrates the methodological approach, emphasizing the key stages of representative data selection, periodogram computation, predominant frequency identification, and pitch shifting, which collectively underpin the functionality and innovative aspects of the proposed system.

3.3.1 Rationale Behind Proposed Method

This method emphasize precision in center frequencies' selection by calculating the periodogram of each audio file and identifying the predominant frequencies. This ensures that these frequencies are more likely to carry important information about the sound.

3.3.2 Custom Center Frequency Calculation

The Mel filter bank center frequencies are found from a subset of a dataset. We initiate the process by calculating the periodogram of each audio file in the subset (Equation 1).

$$P(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-j2\pi f n} \right|^2 \tag{1}$$

After calculating the periodogram for each audio file in the subset, we identify the index of the highest peak within each periodogram to determine the predominant frequencies. These frequencies are then processed through an interval filter to eliminate redundancies, limit the number of bins, and reorder them. This process yields a list of bins tailored to the dataset, effectively capturing essential information across any frequency range.

$$STFT\{x(t)\}(f,\tau) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-j2\pi ft}dt \tag{2}$$

$$|STFT\{x(t)\}(f,\tau)|^2 = \left|\int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-j2\pi ft}dt\right|^2 \tag{3}$$

$$F(u,v) = \frac{2}{\sqrt{MN}}C(u)C(v)\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} f(m,n)\cos\left[\frac{\pi u(2m+1)}{2M}\right]\cos\left[\frac{\pi v(2n+1)}{2N}\right] \tag{4}$$

$$f(i) = \begin{cases} \frac{i-\text{bin\_prev}}{\text{bin\_center}-\text{bin\_prev}} & \text{for bin\_prev} \leq i < \text{bin\_center} \\ \frac{\text{bin\_next}-i}{\text{bin\_next}-\text{bin\_center}} & \text{for bin\_center} \leq i < \text{bin\_next} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

To utilize these custom center frequencies, we devise a bespoke Mel Frequency Cepstral Coefficients (MFCC) function. This function allows users to specify their desired bins, adhering to the overall MFCC methodology. Initially, the librosa library's short-time Fourier transform (STFT) function applies windowing to the signal and calculates its frequency response (as per Equation 2). Subsequently, we square the absolute value of the STFT output (Equation 3), aligning with the third step of the standard MFCC procedure. Next, we generate triangular filters (Equation 5), placing them according to our custom list of center frequencies determined earlier. The application of a logarithm base ten brings us to the process's final stage. We then perform a cosine transform (Equation 4) to extract the 13 commonly used MFCC coefficients, forming a matrix that concludes the preprocessing steps. The resultant matrix, uniquely tailored to the dataset's frequency range, epitomizes our custom approach.

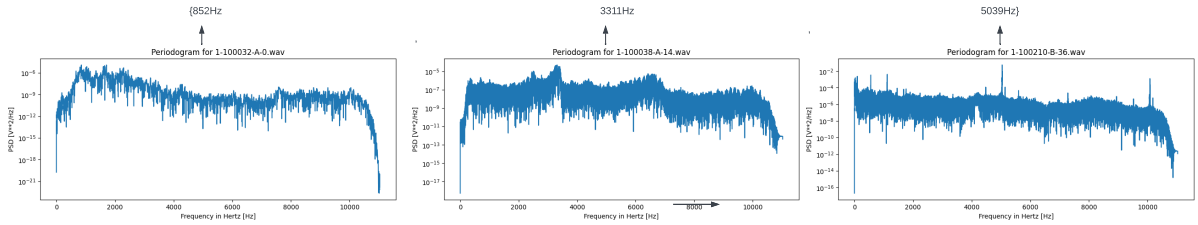


Figure 7: Visualization of Proposed Method

From figure 7, we initially compute the periodogram for each waveform to ascertain the spectrum's power density over various frequencies. After this computation, we undertake the task of identifying the predominant frequencies within each periodogram. This is accomplished by first locating the peaks within the power spectral density data, where we select a subset of the highest peaks by sorting the power density values and retaining the top 500 for further analysis. These peaks are then mapped back to their corresponding frequencies. Following the identification of peak frequencies, our methodology involves sorting these frequencies based on their power density values, ensuring that we prioritize frequencies with higher power densities for subsequent filtering. This prioritization is crucial for isolating the most significant frequencies from the spectrum.

To further refine our frequency selection, we use a filtering method that ensures there's a minimum gap of 100 Hz between any two frequencies in our final list. This is crucial for avoiding frequency overlap and for capturing a wide range of significant frequencies across the spectrum. However, if we struggle to find enough frequencies that are 100 Hz apart, we gradually reduce this gap by 10 Hz at a time and try again. This flexible strategy helps us find the right balance between having a sufficient number of frequencies and keeping them adequately spaced.

We repeat this adjustment process until we have a list of 20 distinct frequencies that adhere to our spacing criteria. Once identified, these frequencies are organized to simplify further examination or comparison. The result is a carefully chosen list of 20 key frequencies,

each selected from the wider spectrum and spaced according to an adjusted gap that ensures a thorough and diverse frequency representation. This method guarantees that we capture an array of frequencies that accurately reflects the waveform's spectral properties.

3.3.3 Pitch Shifting

The pitch shifting process is a sophisticated audio processing technique designed to modify the pitch of an audio signal while maintaining its original duration. This section delves into both the implementation and the conceptual foundation of pitch shifting. Although this step is not mandatory for the overall process, it demonstrates the effectiveness of a custom filter bank at higher frequencies compared to the traditional MFCC approach, highlighting its superiority in accommodating a broader frequency spectrum.

The core concept of pitch shifting involves transposing the audio signal to a higher frequency, ensuring it remains below $f_s/2$ to adhere to the Nyquist Theorem. This adherence guarantees that there will be no loss of information or occurrence of aliasing. This process is crucial for facilitating the application of custom filters across various frequencies, particularly at higher frequencies. It effectively circumvents the conventional challenges faced when implementing MFCC, which typically exhibit a denser distribution at lower frequencies and become sparser at higher frequencies. This strategic approach enables a more balanced and versatile frequency analysis by allowing for the optimized placement of a custom filter bank. Specifically, it facilitates the concentration of filters at higher frequencies, ensuring a denser and more effective analysis in these regions.

3.3.4 Custom MFCC function

The implementation of our Custom MFCCs Function parallels the traditional approach to computing MFCCs, but with a significant adaptation: rather than employing a standard Mel Filter Bank, it incorporates a custom filter bank tailored specifically to our

predefined center frequencies. This modification allows for a more flexible and targeted analysis of audio signals, aligning the extraction process closely with our unique requirements and objectives.

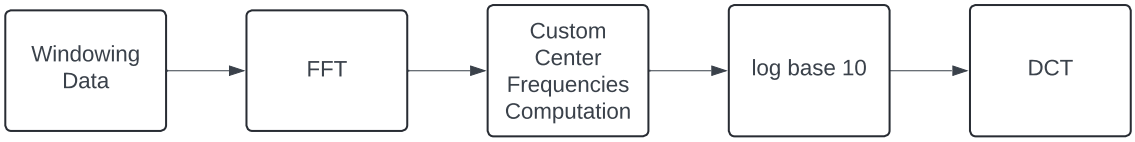


Figure 8: Process of Custom MFCC Computation

In the initial phase of our custom MFCC computation, we apply a Hamming window to the signal, adhering to standard practices in terms of window and overlap sizes. Subsequently, the windowed signal is transformed into the frequency domain via the Fast Fourier Transform (FFT). Following this, a custom MFCC filter bank is employed to further process the data.

Our custom MFCC function, designed for specialized audio processing, accepts an audio signal and various parameters to compute the MFCC features. The function takes in a custom array of center frequencies. This approach allows for the use of custom filter banks that deviate from the standard Mel scale, providing flexibility in capturing frequency characteristics important for specific applications.

For a custom filter bank, it maps the provided center frequencies to FFT bins, constructs triangular filters around these bins, and applies these filters to the power spectrum of the audio signal. This results in a filtered spectrogram tailored to the provided center frequencies. Finally, the function applies the Discrete Cosine Transform (DCT) to the logarithm of the power at each filter, extracting the desired number of MFCC features.

3.4 Machine Learning Model

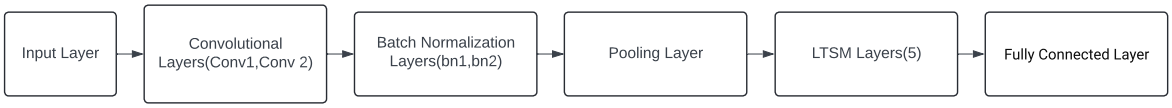


Figure 9: Model Layers illustration

In this paper, we present the integration of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, demonstrating a powerful methodology for efficiently capturing both spectral and temporal features. This section introduces the hybrid CNN-LSTM network, illustrating its effectiveness in audio signal processing.

The network comprises a total of six layers, including a CNN architecture with two convolutional layers (conv1 and conv2), each accompanied by a corresponding batch normalization layer (bn1 and bn2). Additionally, it features a single pooling layer (pool) that is strategically applied twice, following each convolutional layer sequence. In LTSM layers, we have 5 num_layers, there are few reasons that we choose 5 layers. A higher number of LSTM layers enhances the model's capacity for learning complex patterns in sequential data. With 5 layers, the network is designed to capture deep temporal relationships within the input.

Handling Long-term Dependencies: LSTM networks excel at learning long-term dependencies. However, the complexity and span of these dependencies in the data may necessitate multiple layers for effective capture. Utilizing 5 LSTM layers aims to improve the network's capacity to remember and utilize information from much earlier in the sequence, which is crucial for high performance in many sequential modeling tasks.

Enhanced Feature Extraction: Within a hybrid model comprising both CNN and LSTM components, the CNN layers initially extract spatial or spectral features from the

input data. The subsequent LSTM layers, particularly when numerous, are positioned to refine and interpret these features within a temporal context, potentially leading to a more nuanced understanding of the data.

During the forward pass, the input data first traverse the convolutional layers, undergoing a series of transformations that extract relevant spectral features. These features are then temporally analyzed by the LSTM layers, capturing the dynamic characteristics of the audio signal over time. The output of the last LSTM layer is then passed to a fully connected layer, which outputs the final classification results.

4 Experimental Setup and Implementation

4.1 Dataset

This section elaborates on the datasets utilized in our research, including the Respiratory Sound Database, the RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song), and the DonateACry dataset. Each dataset features multiple sound classes. For example, the RAVDESS dataset[26] comprises 14 categories, encompassing a diverse range of human emotional states such as neutral, calm, happy, sad, among others, and contains a total of 7,356 files. This database is enriched by the contributions of 24 professional actors (12 female and 12 male), who perform two lexically-matched statements in a neutral North American accent, providing a rich variety of vocal expressions for analysis.The DonateACry Corpus dataset[28] comprises recordings of baby crying sounds, categorized under various conditions, including belly pain, burping, discomfort, hunger, and tiredness. In respiratory sound database[27], it encompasses 920 annotated recordings, with durations ranging from 10 seconds to 90 seconds. These recordings were collected from 126 patients, culminating in a total of 5.5 hours of audio. This dataset includes 6,898 respiratory cycles, among which 1,864 contain crackles, 886 feature wheezes, and 506 present both crackles and wheezes. The dataset offers a diverse range of audio, from clean respiratory sounds to noisy recordings that mirror real-world conditions. The demographic of patients covered is broad, including children, adults, and the elderly, providing a comprehensive overview of respiratory sounds across different age groups. Additionally, we introduced the "Patient Health Detection using Vocal Audio" dataset[29], which comprises vocal audio files of both healthy individuals and patients diagnosed with specific vocal diseases. The collected audio files encompass two distinct vocal disorders: Laryngocele and Presbylaryngis (formerly referred to as Vox senilis), as well as recordings from individuals without any vocal ailments. This database serves as a crucial resource for distinguishing between normal and pathological

vocal patterns.

4.2 Package Descriptions of Python

In this project, we utilized a selected suite of Python libraries, each chosen for its specific capabilities in facilitating various aspects of our computational experiments, from data handling and preprocessing to model development and optimization.

- **os**: This library is utilized for navigating and manipulating the file system, facilitating efficient data storage and retrieval processes critical to our experimental workflow.
- **numpy (NumPy)**: A cornerstone for scientific computing in Python, NumPy supports high-performance operations on multi-dimensional arrays and matrices, serving as the backbone for numerical computations across our data preprocessing and analysis tasks.
- **torch and its submodules (nn, nn.functional, optim, utils.data)**: PyTorch provides a comprehensive ecosystem for designing, training, and validating deep learning models. Its dynamic computation graph, alongside a rich collection of neural network layers, activation functions, and optimization algorithms, underpins the development of our advanced neural network architectures. The DataLoader and TensorDataset utilities facilitate efficient data management and batching, essential for processing large datasets during model training.
- **librosa**: Specialized for audio processing, librosa enables the extraction of sophisticated features from audio signals, including time-frequency representations, crucial for analyzing the spectral properties of our datasets.
- **scikit-learn**: Offers a wide array of machine learning tools. We employ algorithms like KMeans for clustering, PCA for dimensionality reduction, and utilities for computing

cosine similarity, enhancing our capability to explore and categorize high-dimensional data effectively.

- **matplotlib, seaborn**: These libraries are employed for generating a variety of visualizations to illustrate our findings, from simple line plots to complex, multi-faceted statistical charts, enriching the interpretability of our results.

- **umap**: Facilitates dimensionality reduction for high-dimensional data visualization, enabling us to uncover and interpret complex data structures through simplified, comprehensible representations.

- **scipy**: A broad collection of tools for scientific computing; specific functionalities such as the Discrete Cosine Transform, optimization routines, and statistical models are leveraged to process signals and fit models to our data rigorously.

- **multiprocessing**: Enhances computational efficiency by enabling parallel processing capabilities, significantly accelerating the execution of computationally intensive operations.

5 Results

This section presents the performing of our proposed custom MFCC compared to the standard method. The custom MFCC method is designed with the premise that specific center frequencies can yield a more accurate representation of audio signals for tiny machine learning applications, where model performance and computational efficiency are crucial. In order to test the results of Custom MFCC, we divided the dataset into train and test dataset in a form of 80% of train and 20% test randomly. The research methodology involves designing the custom MFCC function, selecting center frequencies based on signal characteristics, and exploring the theoretical basis for these choices. This is aimed at improving the feature's ability to distinguish between different audio signals.

It also presents how pitch shifting impacts the extraction of features using both the custom and standard MFCC techniques. This aspect is crucial for applications requiring the model to be resilient to variations in signal pitch.

The four datasets chosen for this study vary in audio characteristics to ensure a comprehensive evaluation of the custom MFCC method's effectiveness and robustness compared to the standard approach.

5.1 Performance of Custom MFCC

In the examination of custom and conventional MFCC computation, pitch shifting introduced as an important factor. Since, it helps us to understand altering the pitch of input signal impacts the accuracy and reliability of feature extraction across both methods. Pitch shifting, a common phenomenon in audio processing, can significantly impact the performance of machine learning models, especially in applications where the audio input is subject to variations in tone or frequency.

The Friedman test was conducted to compare the performance of custom Mel-Frequency Cepstral Coefficients (MFCC) and conventional MFCC across datasets. The Friedman test statistic was found to be 18.37, with a corresponding p-value of 0.0104. This result indicates a statistically significant difference in MFCC performance between custom MFCC and conventional MFCC across the datasets at a significance level of $\alpha = 0.05$. To systematically analyze the effect of pitch shifting on both methods, we applied pitch shifting to all datasets.

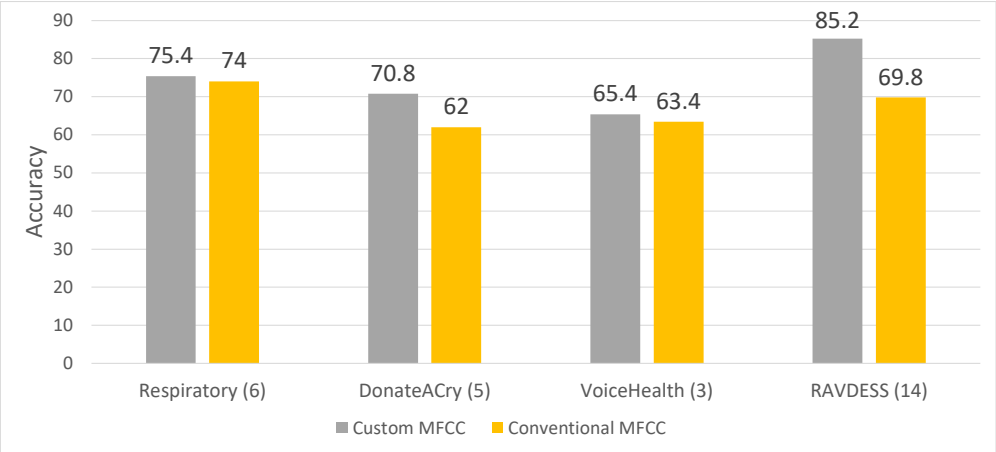
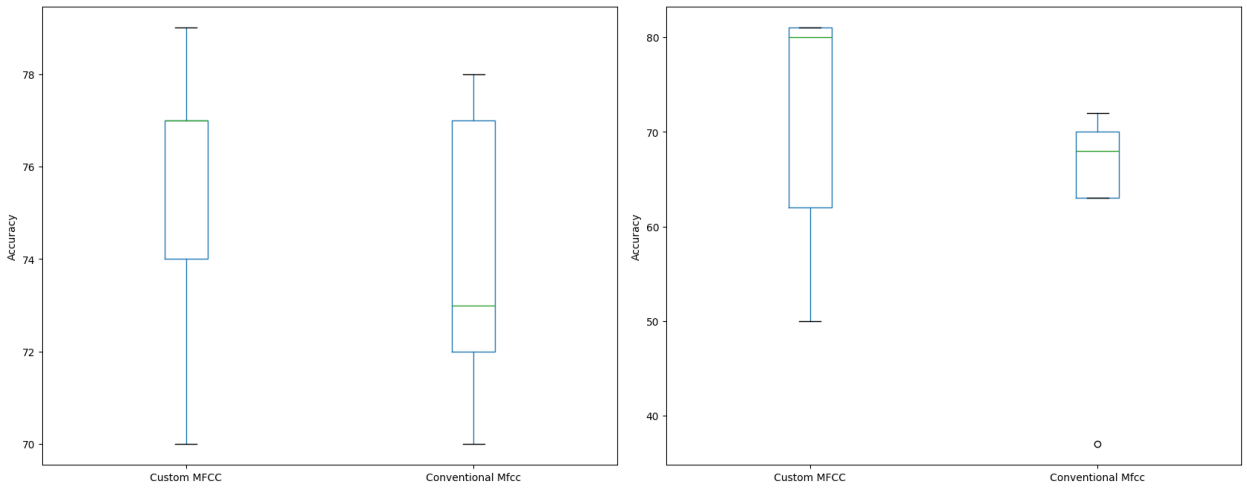


Figure 10: Custom MFCCs with Shift Comparing with conventional MFCC with Shift

From the figure 10, we analyze the performance of two different MFCC methods applied across four datasets: Respiratory, DonateACry, VoiceHealth, and RAVDESS.

The figure compares Custom MFCC with Shift against a Normal shift method. The y-axis is accuracy and x-axis is dataset. Respiratory (6 classes): The Custom MFCC with pitch shifting achieves 75.4%, which basically maintains the same as conventional MFCC of 74%. DonateACry (5 classes): Here, the Custom MFCC method scores 70.8%, whereas the conventional approach has a much lower score of around 62%. VoiceHealth (3 classes): The performance of the Custom MFCC method is 65.4%, outperforming the conventional method's score of around 63.4% by 2%. RAVDESS (14 classes): The Custom MFCC method's score is 85.2%, whereas the conventional method scores around 69.8%. Across all datasets, the Custom MFCC with Shift method slightly outperforms the Normal MFCC with Shift method. This indicates that the custom approach to implementing MFCC with pitch shifting offers a slight advantage over the conventional approach, at least within the context of these datasets. Pitch shifting as a data augmentation technique can help in capturing a wider variety of features by simulating different pitch levels, which might not be present in the original data and cannot be capture by conventional MFCC method. This could explain the improved performance, as the model trained with the custom MFCC with pitch shifting method is likely more robust and generalizes better across varied audio samples.

5.1.1 Respiratory Dataset



Respiratory Dataset DonateACry Dataset

Figure 11: Performance of Custom MFCCs conventional MFCC

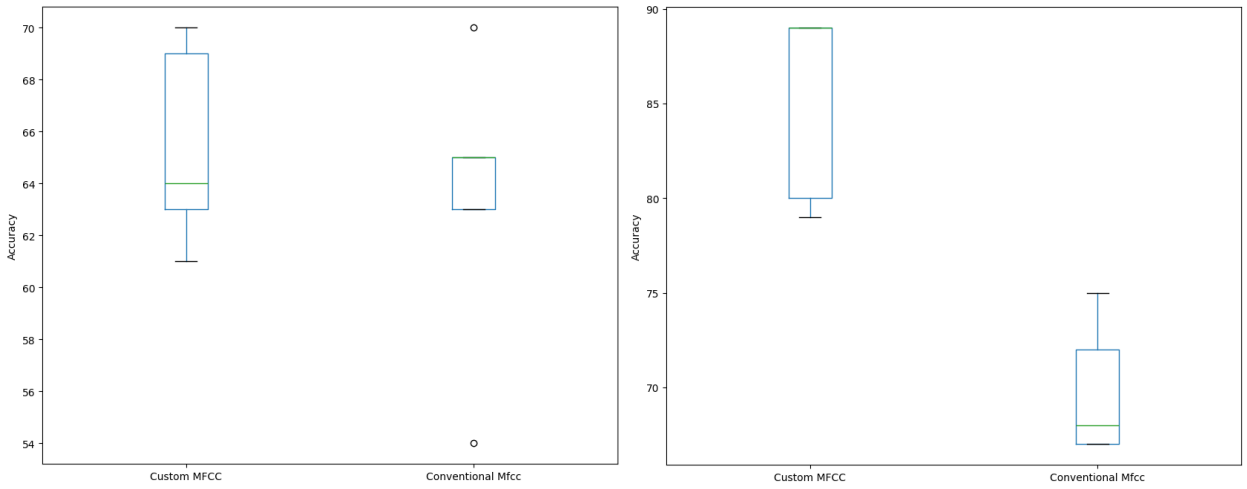
From figure 11 Respiratory Dataset, the diagram is computed using 5 run results from both method. The Custom MFCC pitch shift has a slightly higher median score than the conventional MFCC pitch shift, indicating that it generally scores higher. The interquartile range (IQR), which is the range of the middle 50% of the data, is similar for both methods, suggesting similar variability in scores across the runs. There are no outliers for either method, which implies that all runs are within an expected range without extreme variations. The plot suggests that while both methods perform similarly, the Custom MFCC method with pitch shift has a slight edge in terms of median performance on the Respiratory dataset.

5.1.2 DonateACry Dataset

Figure 11 DonateACry Dataset demonstrates that the Custom MFCC pitch shift method exhibits a broader variability and an extended range of values, highlighted by a notable escalation towards higher values within the dataset. This suggests that while Custom

pitch shifts vary more extensively, they align with a superior accuracy score. Conversely, the conventional MFCC pitch shift dataset is characterized by a more constrained distribution, indicating a more uniform collection of pitch shift values, albeit associated with lower accuracy scores.

5.1.3 VoiceHealth Dataset



VoiceHealth Dataset RAVDESS Dataset

Figure 12: Performance of Custom MFCCs conventional MFCC

From figure 12 VoiceHealth, it suggests that the Custom MFCC pitch shift dataset, although showing a slightly wider range of values compared to the conventional MFCC pitch shift, indicates a balanced spread around the median, reflecting a moderate level of variability. Conversely, the conventional MFCC pitch shift presents a highly concentrated distribution, pointing towards a consistent performance with minimal deviation among its values. It means that the Custom MFCC score a higher accuracy, although it's wider comparing to conventional MFCC, which is denser but score a lower accuracy.

5.1.4 RAVDESS Dataset

From figure 12 RAVDESS Dataset, it suggests that Custom MFCC pitch shift method propensity towards higher values with a modest range, indicating a focused yet slightly varied performance. Conversely, the conventional MFCC pitch shift presents a more uniform distribution, characterized by its narrow spread and lower pitch shift values, suggesting a consistent performance with less variability.

5.2 Effects of pitch shifting

In this subsection, we focus on the comparison between custom MFCC computation and the conventional approach, specifically excluding pitch shifting. Pitch shifting, while significant in audio processing for its effects on model performance due to tone or frequency variations, is not considered in this analysis. Here, the examination centers on understanding the impact of excluding pitch shifting on the accuracy and reliability of feature extraction within both custom and conventional MFCC methodologies. This approach allows us to isolate and evaluate the core differences and efficiencies of each method in processing audio signals without the influence of pitch variations. By understanding these aspects, the section aims to shed light on the fundamental characteristics of custom MFCC versus conventional MFCC computation and their implications for machine learning algorithms in audio data analysis, directing towards the advancement of audio processing techniques that do not rely on pitch adjustment.

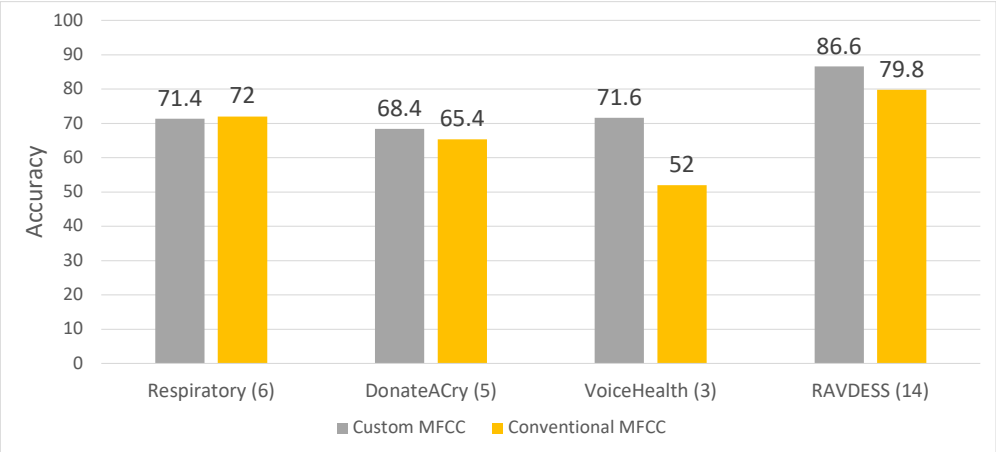
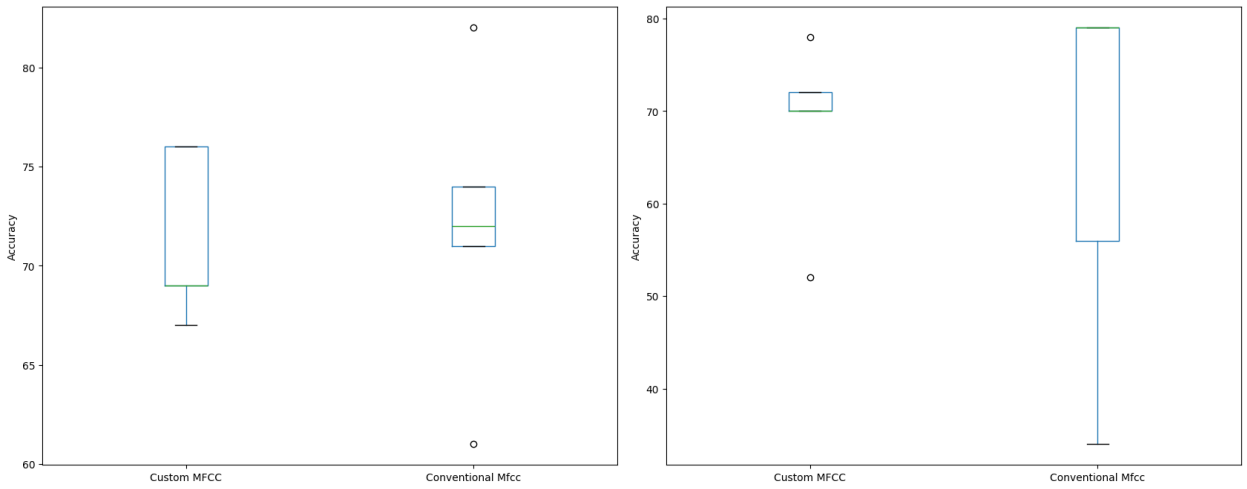


Figure 13: Custom MFCCs Comparing with conventionalMFCC

Figure 13 presents a comparison between Custom MFCC and conventional MFCC calculations across four distinct datasets. In the Respiratory dataset, accuracy levels are nearly identical, with conventional MFCC at 72% and Custom MFCC slightly lower at 71.4%. The DonateACry dataset demonstrates a 3% improvement in accuracy with Custom MFCC over the conventional method. The VoiceHealth dataset reveals a significant enhancement in performance with Custom MFCC, attributable to two main factors: firstly, the dataset's complex vocal sounds, crucial for health diagnostics, contain subtle variations more effectively captured by the Custom MFCC, which is likely optimized to extract nuanced health-related features. Secondly, the dataset's limited number of classes (three) necessitates high precision to accurately differentiate between closely related health sound categories. Lastly, the RAVDESS dataset shows a 6.8% increase in accuracy with Custom MFCC compared to conventional MFCC, highlighting the effectiveness of the custom method across diverse audio analysis tasks."

5.2.1 Respiratory Dataset



Respiratory Dataset DonateACry Dataset

Figure 14: Performance of Custom MFCCs conventional MFCC without pitch shift

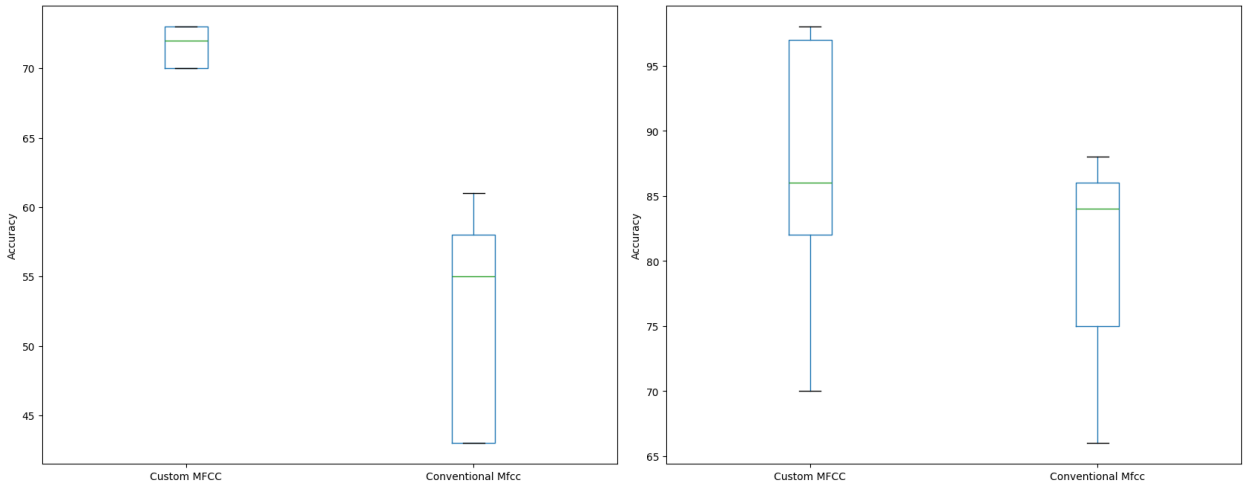
From figure 14 Respiratory Dataset, the Custom MFCC method demonstrates a wider spread in its performance outcomes, suggesting a potential for both higher and lower results but with a tendency towards the median being at the lower end of the range. On the other hand, the conventional MFCC method shows a more consistent set of outcomes, with a narrower range of variability, indicating a steady performance across different metrics. This consistency could imply a more predictable and reliable performance, albeit with less potential for reaching the higher outcomes observed in the Custom MFCC method.

5.2.2 DonateACry Dataset

From figure 12 DonateACry Dataset, these results indicate that the Custom MFCC non pitch shift method has a tighter quartile spread and a smaller range of values considered non-outliers, suggesting a more concentrated set of values around the median. The conventional MFCC non pitch shift method shows a wider quartile spread and a broader range of non-outlier values, indicating more variability in the data. The presence of the same value

for the median and Q3 in the conventional dataset suggests a skewed distribution with a cluster of high values.

5.2.3 VocieHealth Dataset



VocieHealth Dataset RAVDESS Dataset

Figure 15: Performance of Custom MFCCs conventional MFCC without pitch shift

From figure 14 VocieHealth Dataset, these results for the reduced datasets indicate a narrower range of values for both datasets compared to the original ones, with a more concentrated set of values around the median for the Custom MFCC non pitch shift. The conventional MFCC non pitch shift dataset shows a significant range between the whiskers, indicating variability but with a more consistent spread around the median than in the original dataset. The reduced Custom MFCC pitch shift dataset shows a very tight quartile spread, suggesting very little variation among most of its values.

5.2.4 RAVDESS Dataset

From 15 RAVDESS Dataset, it shows that the Custom MFCC method displays a greater potential for high performance but with a risk of more significant variability. In

contrast, the conventional MFCC method offers a more consistent set of outcomes, with less extreme highs and lows. This analysis is crucial for decision-making in applications where the choice between potential high performance and consistency needs to be balanced based on the specific requirements and tolerances of the task at hand.

5.3 Analysis between both methods including pitch shift variants

Comparing the Custom MFCC and conventional MFCC methods, including their pitch shift variants, reveals distinct patterns in terms of performance variability, consistency, and potential accuracy.

Performance Variability: The Custom MFCC method, across the datasets analyzed, typically shows a wider range of performance metrics, as indicated by broader interquartile ranges (IQR) and whisker spans. This suggests that the Custom MFCC method can achieve higher performance peaks but also exhibits a greater variability. In contrast, the conventional MFCC method demonstrates a narrower range of performance, suggesting more consistent outcomes across different scenarios but with a potential limitation in reaching the high scores observed in the Custom MFCC method.

Potential Accuracy and Performance Peaks: When examining the potential for achieving higher performance metrics, the Custom MFCC method often displays the capacity for superior accuracy, as seen in the datasets where higher maximum values were recorded. This suggests that under the right conditions or with the appropriate tuning, the Custom MFCC method could be optimized to achieve exceptional results, outperforming the conventional method.

6 limitation

The limitations of this method can adversely affect its performance, accuracy, and cost. By computing the predominant frequencies of each audio wave and applying them as templates across the entire dataset, the method may become computationally expensive. This is particularly evident as the balance between accuracy and cost shifts; achieving higher accuracy results demands significantly more computational resources. Additionally, a partial dataset may not adequately represent certain sound files, compromising effectiveness. Moreover, the approach of computing representative frequencies for each audio file carries the risk of including noisy files. This could result in a low Signal-to-Noise Ratio (SNR), thereby negatively impacting the accuracy of the predominant frequencies identified.

7 Conclusion

In this paper, we explore an innovative approach to enhancing audio classification by customizing MFCC processing, particularly for applications in health monitoring and technology. Our methodology emphasize the computation of custom MFCCs that better represents specific characteristics of the datasets in question, compared to conventional MFCC computation methods. This customization is based on calculating periodograms to identify predominant frequencies and creating a tailored filter bank, thereby overcoming the limitations of the logarithmic scale distribution in conventional MFCCs.

The results of our experiments across various human-related audio datasets demonstrated a significant improvement in classification accuracy when utilizing pitch-shifted audio samples. This underscores the custom MFCC's superior capability in handling human sound variations, thereby enhancing the performance of audio classification systems.

However, our methodology is not without limitations. The computation of custom

MFCCs can be more computationally expensive than conventional methods. Additionally, our approach's effectiveness is partially dependent on the selection of representative frequencies, which may not always perfectly capture the nuances of certain audio files, especially in the presence of noise.

In conclusion, this thesis introduces a method that improves audio classification systems using custom MFCC processing. This method shows promise for enhancing sound analysis technologies, including health monitoring and environmental sound classification. Future research may focus on making custom MFCC computation more efficient and applicable to more audio analysis areas.

References

[1] W. Han, Z. Zhang, Y. Zhang, J. Yu, C.-C. Chiu, J. Qin, A. Gulati, R. Pang, and Y. Wu, "Contextnet: Improving convolutional neural networks for automatic speech recognition with global context," 2020.

[2] M. Orken, D. Oralbekova, K. Alimhan, T. Tolganay, and M. Othman, "A study of transformer-based end-to-end speech recognition system for kazakh language," *Scientific Reports*, vol. 12, 05 2022.

[3] Y. Zhang, "Music recommendation system and recommendation model based on convolutional neural network," *Mobile Information Systems*, vol. 2022, pp. 1–14, 05 2022.

[4] H. Chu, Y. Zhang, and H.-C. Chiang, "A cnn sound classification mechanism using data augmentation," *Sensors*, vol. 23, pp. 6972–6972, 08 2023.

[5] M. Green and D. Murphy, "Environmental sound monitoring using machine learning on mobile devices," *Applied Acoustics*, vol. 159, p. 107041, 02 2020.

[6] D. Goyal and B. S. Pabla, "The vibration monitoring methods and signal processing techniques for structural health monitoring: A review," *Archives of Computational Methods in Engineering*, vol. 23, pp. 585–594, 03 2015.

[7] T. Nishida, K. Dohi, T. Endo, M. Yamamoto, and Y. Kawaguchi, "Anomalous sound detection based on machine activity detection," in *2022 30th European Signal Processing Conference (EUSIPCO)*, pp. 269–273, 2022.

[8] Y. Wang, Y. Zheng, Y. Zhang, Y. Xie, S. Xu, Y. Hu, and L. He, "Unsupervised anomalous sound detection for machine condition monitoring using classification-based methods," *Applied Sciences*, vol. 11, p. 11128, 11 2021.

[9] H. Harb and L. Chen, "Gender identification using a general audio classifier," in *2003 International Conference on Multimedia and Expo. ICME '03. Proceedings (Cat. No.03TH8698)*, vol. 2, pp. II–733, 2003.

[10] F. Rong, "Audio classification method based on machine learning," in *2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, pp. 81–84, 2016.

[11] B. Vimal, M. Surya, Darshan, V. Sridhar, and A. Ashok, "Mfcc based audio classification using machine learning," pp. 1–4, 07 2021.

[12] M. Slaney, "Auditory toolbox," 12 1993.

[13] S. Suksri, "Speech recognition using mfcc," 09 2015.

[14] J. T. Geiger, B. Schuller, and G. Rigoll, "Large-scale audio feature extraction and svm for acoustic scene classification," in *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 1–4, 2013.

[15] N. Esfandian, F. Razzazi, and A. Behrad, "A clustering based feature selection method in spectro-temporal domain for speech recognition," *Engineering Applications of Artificial Intelligence*, vol. 25, pp. 1194–1202, 09 2012.

[16] X. Xie, H. Cai, C. Li, Y. Wu, and F. Ding, "A voice disease detection method based on mfccs and shallow cnn," *Journal of Voice: Official Journal of the Voice Foundation*, pp. S0892–1997(23)00301–6, 10 2023.

[17] A. Suppa, G. Costantini, F. Asci, P. Di Leo, M. S. Al-Wardat, G. Di Lazzaro, S. Scalise, A. Pisani, and G. Saggio, "Voice in parkinson's disease: A machine learning study," *Frontiers in Neurology*, vol. 13, 02 2022.

[18] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM International Conference on Multimedia*, MM '10, (New York, NY, USA), p. 1459–1462, Association for Computing Machinery, 2010.

[19] M. Jung and S. Chi, "Human activity classification based on sound recognition and residual convolutional neural network," *Automation in Construction*, vol. 114, p. 103177, 06 2020.

[20] L. Verde, G. De Pietro, and G. Sannino, "Voice disorder identification by using machine learning techniques," *IEEE Access*, vol. 6, pp. 16246–16255, 2018.

[21] Y. Zayed, A. Hasasneh, and C. Tadj, "Infant cry signal diagnostic system using deep learning and fused features," *Diagnostics*, vol. 13, no. 12, 2023.

[22] X. Peng, H. Xu, J. Liu, J. Wang, and C. He, "Voice disorder classification using convolutional neural network based on deep transfer learning," vol. 13, 05 2023.

[23] M. Nicolini, F. Simonetta, and S. Ntalampiras, "Lightweight audio-based human activity classification using transfer learning," *Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods*, 2023.

[24] S. Abbas, S. Ojo, A. Al Hejaili, G. A. Sampedro, A. Almadhor, M. M. Zaidi, and N. Kryvinska, "Artificial intelligence framework for heart disease classification from audio signals," *Scientific Reports*, vol. 14, p. 3123, 02 2024.

[25] A. H. Sabry, O. I. Dallal Bashi, N. Nik Ali, and Y. Mahmood Al Kubaisi, "Lung disease recognition methods using audio-based analysis with machine learning," *Heliyon*, vol. 10, p. e26218, 02 2024.

[26] S. R. Livingstone and F. A. Russo, "The ryerson audio-visual database of emotional speech and song (ravdess): A dynamic, multimodal set of facial and vocal expressions in north american english," *PLOS ONE*, vol. 13, p. e0196391, 05 2018.

[27] B. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jacome, A. Marques, R. P. Paiva, I. Chouvarda, P. Carvalho, and

N. Maglaveras, "A respiratory sound database for the development of automated classification," in *Proceedings of the International Conference on Digital Health*, pp. 33–37, 2017.

[28] G. Veres, "gveres/donateacry-corpus," 03 2024.

[29] T. Ozseven, "Infant cry classification by using different deep neural network models and hand-crafted features," *Biomedical Signal Processing and Control*, vol. 83, p. 104648, 2023.