IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Mel Frequency Cepstral Coefficient and its applications: A Review

Zrar Kh. Abdul[1,2], and Abdulbasit K. Al-Talabani[3]

1 Department of Information System Engineering, Erbil Technical Engineering College, Erbil Polytechnic University, Erbil, Kurdistan Region, Iraq
2 Department of Applied Computer Science, Charmo University, Iraq
3 Department of Software Engineering, Faculty of Engineering, Koya KOY45, Kurdistan Region – F.R. Iraq

Corresponding author: Zrar Kh. Abdul (eng.zrar1394@gmail.com).

ABSTRACT Feature extraction and representation has significant impact on the performance of any machine learning method. Mel Frequency Cepstrum Coefficient (MFCC) is designed to model features of audio signal and is widely used in various fields. This paper aims to review the applications that the MFCC is used for in addition to some issues that facing the MFCC computation and its impact on the model performance. These issues include the use of MFCC for non-acoustic signals, adopting the MFCC alone or combining it with other features, the use of time series versus global representation of the MFCC, following the standard form of the MFCC computation versus modifying its parameters, and supplying the traditional machine learning methods versus the deep learning methods.
.

INDEX TERMS MFCC; Speech analysis; Cepstrum analysis; feature extraction, indusial analysis

1. INTRODUCTION

The complexity of any system is a critical issue, though reducing it has been a promising area for researchers. One of the basic demands for any machine learning application is to have an enormous size dataset in terms of number of samples and variables. Nowadays, the size of the collected dataset tends to be larger such that it leads to the increase of the adopted system complexity. For this purpose, many approaches have been developed to reduce dimensionality[1]. Dimensional reduction helps in data compression since it reduces storage space and computation time. It is also a way to eliminate redundancies from the dataset. Dimensional reduction can be conducted in using feature extraction and feature selection techniques[2].

The feature extraction is the process of extracting and tackling hidden information in the raw data signal[3]. Data can be more manageable by applying feature extraction because it removes all ineffective features from the data without losing any important or relevant data. Moreover, feature extraction techniques help to develop a system with less machine's efforts and increase the speed of learning and generalization steps in the machine learning process[4], [5].

Nowadays, many feature extraction techniques are available in a variety of fields based on the characteristics of the raw data. In most of the fields, finding harmonics and sidebands of signal in both time and frequency domain are important to any pattern recognition system. Power spectrum using Fast Fourier Transform (FFT) is used to capture the harmonics and sidebands of the signal in the time domain.

While cepstrum; such as Mel Frequency Cepstrum Coefficient (MFCC), Gamma Tone Cepstrum Coefficient (GTCC), is capable to extract harmonics and sidebands of the spectrum version of the signal[6].

Although MFCC is used to extract feature from input data of various fields, it faces many challenges, which have not been covered comprehensively in the literature. The purpose of this paper is to conduct a comprehensive review of MFCC and its applications including speech recognition, speaker recognition, emotion recognition, bearing fault detection, gear fault detection, Electrocardiogram (ECG) and Electroencephalogram (EEG) classification. Additionally, the paper targets answering the following questions regarding the use of MFCC:

1. How wide is MFCC applied to other fields rather than acoustic signal?
2. Is the MFCC well-tuned to fit the demand of various applications?
3. Since the MFCC is computed in short term signal, has it been used more in a time-series or in a global representation?
4. Is MFCC mostly combined with other features, or it is used alone?
5. Does MFCC feed deep learning more or traditional machine learning methods?
6. What are the most well-known used windows in the data framing process?
7. Which classifier is used more frequently with MCFF and what could be the reason?

IEEE *Access*
Multidisciplinary ⫶ Rapid Review ⫶ Open Access Journal

The rest of this paper is organized as follows; the computation of MFCC is explained in section 2. Section 3 introduces MFCC applications. Followed by a summarized analysis presentation in section 4, and finally the conclusion is drowned in section 5.

2. MFCC IMPLEMENTATION

MFCC is one of the commonly used features that has been used in a variety of applications especially in voice signal processing such as speaker recognition, voice recognition, and gender identification [6]. The MFCC can be calculated by conducting five consecutive processes, namely signal framing, computing of the power spectrum, applying a Mel filter bank to the obtained power spectra, calculating the logarithm values of all filter banks, and finally applying the DCT. Figure (1) illustrates the processes of MFCC computation [7].
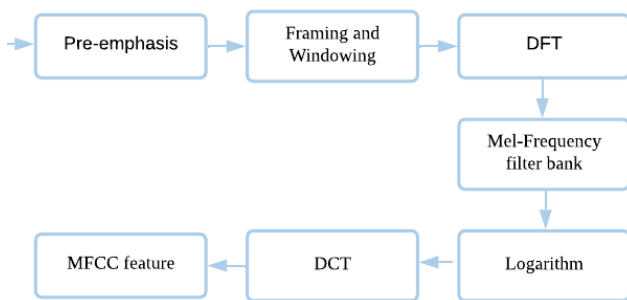

Figure 1: MFCC framework

2.1. Pre-emphasis

Pre-emphasis is one of the common pre-processing practices in signal processing area which is used to compensate the high frequency of the signal that was suppressed during the signal production. Pre-emphasis is the very first step during the MFCC adaptation, which can be adopted by simply applying a high-pass filter with a setting of [1, -0.97]. The filtering process alters energy distribution across frequencies, as well as the overall energy level[8].

2.2. Signal framing and windowing

The idea behind splitting signals into distinct "frames" is to break down the raw data signal into frames where the signal tends to be more stationary. For stable acoustic characteristics, speech needs to be examined over a sufficiently short period of time. Regarding the speech signal, the period of 20-30ms time is reported to be a Quasi-Stationary Segment (QSS) since the time between two glottal closure is shown to be around 20ms. However, vowel voices are reported to be captured in 40ms- 80ms [9]. Hence, short-term spectral measurements are typically carried out over 20 ms windows, and each frame is overlapped by 10 ms with the next one. Overlaps of frames by 10 ms enables the temporal characteristics the speech signal to be tracked. With overlapping of frames speech, sound representation would be approximately centered at some frame.

On each frame, a window is applied to narrow the signal towards the border of the frame. In general, Hanning and Hamming windows[10] are among the most well-known nominees. These windows can enhance harmonics, smooth edges, and diminish edge effect while taking a DFT on the signal. Figure (2) illustrates the rectangular Hamming and Hanning windows in both time and frequency domains.
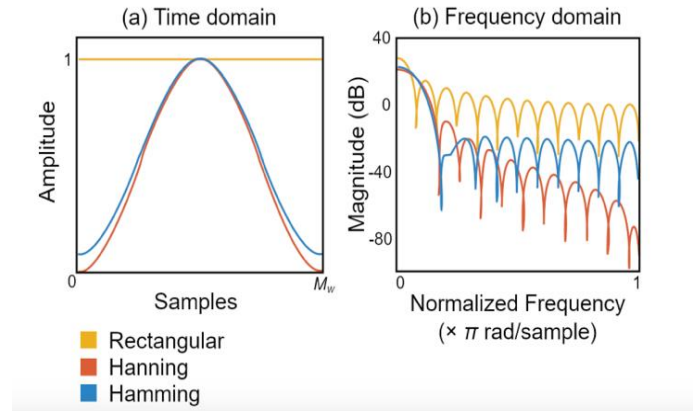

Figure 2: The Rectangular Hanning and Hamming windows in the time and frequency domains[11].

2.3. Power spectrum

A power spectrum can be described as the distribution of the power of the frequency components that composes the signal[12]. Traditionally, Discrete Fourier Transform (DFT) is utilized to compute the power spectrum. The power spectrum of each of the obtained frames must be determined based on the below equation (1):

$$X(k) = \sum_{n=0}^{N-1} x(n)\, e^{-\frac{2\pi j n K}{N}} \qquad k = 1,2,3\dots N-1 \quad (1)$$

Where $x(n)$ is discrete signal and N is the length of the signal.

2.4. Mel filter bank

The Mel band-pass filter is a bank of filters, which is constructed based on pitch perception. The Mel filter was originally developed for speech analysis and like human ear perceiving of speech, it targets extracting non-linear representation of the speech signal. The convention Mel filter-bank is constructed of 40 triangular filters [13]. The transfer function (TF) of each of the m-th filter can be computed via equation (2),

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \dfrac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \le k < f(m) \\ 1 & k = f(m) \\ \dfrac{f(m+1) - k}{f(m+1) - f(m)} & f(m) < k \le f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad (2)$$
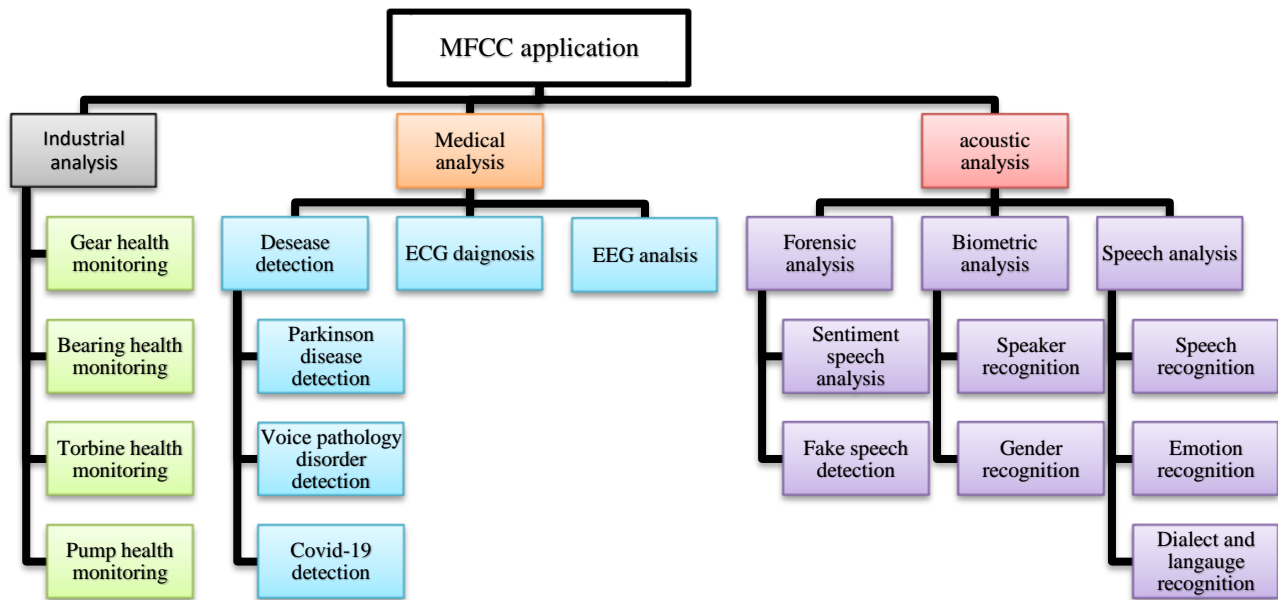
Figure 3: Common application of MFCC

Where, $f\ (m)$ is the centre frequency of the triangular filter and $\sum_{m}^{M-1} H_m(k) = 1$.

The Mel scale to the response frequency and vice versa is computed by equations (3) and (4)

$$m = 2595log_{10}\left(1 + \frac{f}{700}\right) \qquad (3)$$

$$f = 700\left(10^{m/2595} - 1\right) \qquad (4)$$

2.5. Discrete Cosine Transforms (DCT)

A Discrete Cosine transform (DCT) expresses a finite sequence of data points regarding a summation of cosine functions oscillating at different frequencies. The DCT was introduced by Nasir Ahmed in 1972. In the MFCC process, the DCT is applied on the Mel filter bank to select most accelerative coefficients or to separate the relationship in the log spectral magnitudes from the filterbank [14]. The DCT is computed by the below equation (5)

$$X(k) = \sum_{n=0}^{N-1} x_n * COS\left(2\pi jnk/N\right), \qquad k = 1,2,3 \dots N-1 \quad (5)$$

Where x_n is discrete signal and N is the length of the signal.

3. MFCC APPLICATIONS

Nowadays, many huge datasets are available that have been collected in various fields. It is not applicable to deal with the whole of these datasets, which are used to develop automatic systems [15]. Mel Frequency Cepstrum Coefficient (MFCC) is a framework feature extraction technique that has been

adopted in many areas and frequently reported to be useful for various applications. In this section a set of these applications are presented (see figure 3).

3.1. MFCC in Acoustic analysis

3.1.1. Speech analysis

Speech signal analysis aims to find out more informative, compact, and relevant knowledge than the speech signal raw data itself. Vocal tract features (also named as segmental, spectral or system features) [16] are one of the well-known representations of speech analysis. When the process of speaking starts, the air is traveling out from the lungs, the air movements within the vocal tract create a unique version of the sound. Vocal tract features are well reflected in the frequency domain of the speech signal. There are various developed techniques to extract vocal tract features, one of which is the MFCC. The MFCC can capture the vocal tract features, since it extracts human ear information, which has a response of a non-linear scale instead of a linear scale[17]. Among the most well-known applications that utilize MFCC of the speech signal are emotion recognition in speech, language and dialect recognition and speech recognition as shown in the following subsections.

.

3.1.1.1. Automatic speech recognition:

The ASR model targets the recognition of four main categories, which are isolated word, connected word, continuous speech, and spontaneous speech [18]. Regarding the isolated word speech recognition, the speaker needs to pause briefly between words. While connected work recognition aims to recognize words that are separated with a minimum pause. On the other hand, a continuous speech

recognition system does not require pause between words as the speech is continuous [19]. The continuous speech recognition model is dealing with an almost natural sound. However, spontaneous speech can handle natural speech in various languages. One of the shared challenges in all the models is providing a generalized and robust feature of the speech signal to achieve high recognition accuracy.

The MFCC feature is one of the well-known features especially for isolated word speech recognition for instance the MFCC feature was extracted from speech signals of spoken words (isolated) and then classified via Support Vector Machine (SVM) and Maximum Likelihood Classifier [20]. Dhingra et al. proposed a model to isolate 10 digits through their speech signal. The MFCC was extracted from the spoken words and identified based on a similarity measurement which was measured by Dynamic Time Warping (DTW) [21]. Omer et al. used MFCC with two other spectral based features to feed a pairwise SVM classifier and the model was used for isolating uttered Kurdish digits (0-9) recognition systems[22]. Author in[23] used MFCC to feed an Artificial Neural Network (ANN) classifier to identify the isolated Arabic words.

In 2019, NASSIF et al. reviewed 174 papers which were developed ASR models based on deep learning models where all the deep learning models were fed by extracting specific features. Moreover, all the papers (174) have been published between 2006 to 2018. They found that most of the researchers, which was about %69, still use MFCCs as a feature to feed the machine learning models [24].

Despite these usages of MFCC, there is a limitation of this feature for speech recognition system due to the fixed window size (20ms-30ms) and this limitation leads to two problems. Firstly, the obtained frequency resolution for the Quasi-Stationary Segment (QSS) longer than 20 ms is quite low compared to what could be achieved using longer analysis windows. Secondly, the analysis window can span the transition between two QSSs and the MFCC vectors extracted from such transition segments do not provide the information about a single unique (stationary) class which may lead to poor discrimination in a pattern recognition problem [9]. Table (1) summarizes some research, where the MFCC is involved in speech recognition application in the past decade.

3.1.1.2. Speech Emotion Recognition:

Speech Emotion Recognition (SER) is the process of identifying the emotion of the human from the speech signal. Based on the existing literature, it is well understood that MFCC features for SER can reach a high recognition rate compared to other cepstrum features because it is a short-term spectral based feature that leads to extract a rich amount of information from speech signals [38]. Consequently, the MFCC is adopted frequently for SER by the researchers for instance: Aouani and Ayed examined two set of features that includes MFCCs. Features were examined using 39 MFCCs and 65 MFCCs together with the SVM classifier, which was used to evaluate both sets. The finding of this paper shows that the first set which was 39 MFCCs, can outperform 65

MFCCs[39]. On the other hand, the related information to the emotion is reported to be sparse and distributed in various features [40]. Consequently, researchers integrated MFCCs with other sort of features, for example, MFCCs and Gamma-Tone Cepstral Coefficients (GTCCs) features were fused together for SER and fed either to the GMM classifier [41] or ESN model [42]. Improvement in emotion recognition has also been noticed by hybridizing some features including formants, pitch, zero crossing, MFCC and its statistical parameters[43]. Aouani et al. studied a model for emotion recognition based on the SVM where the SVM was fed by 42-dimensional vector which consisted of 39 coefficients of MFCC, a Teager Energy Operator, a Zero Crossing Rate, and a Harmonic to Noise Rate. The obtained feature (42-dimensional vector) was reduced by Auto-Encoder, the results confirmed that the reduced dimension representation effectively improved the SER accuracy[44]. However, in another study, six basic emotional states were classified based on GMM, which was fed by two sets of features, namely Sub-band based Cepstral Parameter and MFCC features. Here, the Sub-band based Cepstral Parameter features is reported to outperform the MFCC features[45]. Table (2) provides more details about other research that used MFCC for human emotion and music emotion recognition.

3.1.1.3. Language and Dialect recognition:

Language recognition system aims to recognize the language of the speaker, which can be developed based on the machine learning model. Dialect recognition is quite like language recognition. However, the dialect recognition focuses on categorizing the dialects or accents rather than categorizing the languages and consequently, more challenging due to the linguistic similarities among dialects. Dialect recognition is an important application, and it has been considered as one of the very first steps to speech recognition. This may be the reason for the plenty of articles that have been published. Various features for both language and dialect recognition are available in the literature such as MFCC (see table (3)), singular values decomposition and Linear Predictive Codes (LPC).

Koolagudi et al. extracted several MFCC sets, which were different in dimension (6, 8, 13, 19, 21, 29 and 35), from the speech signal for identifying a language such as Bangla, Assamese, Guajarati, Hindi, Nepali, Kashmiri, Malayalam, Marathi, Rajasthani, Odia, Punjabi, Tamil, Telugu, Kannada, and Urdu. Among these MFCC sets (6, 8, 13, 19, 21, 29 and 35) nineteen dimensions outperformed the other sets [48].

Warohma et al. developed a model for Indonesian dialect recognition based on the multilayer perceptron. The multilayer perceptron was fed by obtaining MFCC from the speech signal [49]. Tawaqal and Suyanto claimed that the use of the traditional machine learning approach gives a low accuracy for dialect recognition, and they proposed a model for increasing the performance of Indonesian dialects recognition system based on deep recurrent neural network (DRNN) where the DRNN was trained by MFCC feature

[50]. Mansour et al. investigates the usage of MFCC and LPC for language identification, namely Arabic, English and French. Both features were evaluated based on the Artificial Neural Network (ANN) and as a result, MFCC shows better performance to identify these languages compared to LPC[51].

The MFCC feature was extracted by conducting MFCC steps and fed to the GMM approach for identified dialects of Himachal Pradesh. The result shows high performance of MFCC for identification of some dialects [52]. However, Al-Talabani et al [53], evaluated three features including the MFCC, Local binary pattern (LBP), and LPC separately and as a merged form. The adopted features used to feed the SVM model to identify Kurdish dialects and distinguish between some languages including Kurdish, Arabic, Turkish, and Persian. The result shows that the LPC has better performance than MFCC and a fusion feature LBP-LPC outperformed the rest of the feature's form.

Based on the existing literature, the MFCC features do not always show capability to outperform other well-known features such as LPC. This may be due to the language and dialect properties in addition to the nature of collected data. However, the studies in the literature show that it is one of the informative features for language and dialect recognition.

3.1.2. Biometric application

Biometric application is an application which allows to either confirm or determine the identity of a certain characteristic of an individual. The aim of these applications is to confirm that the provided services are accessed only by a legitimate user. Biometric applications can be classified into two categories based on their task such as verification and identification application, where the verification application validates a person's identity by comparing the captured biometric information with her/his own biometric template which is stored in the database and the relationship between the objective and the database is called one to one. While, identification application is called one to many as the application tries to identify an individual by searching the templates of all the users in the database [22]. Based on the existing literature, the MFCC has been used in two biometric applications, speaker recognition and gender recognition over the phone call.

3.1.2.1. Speaker recognition

Speaker recognition is the process of automatically recognizing the speaker based on their speech signal and can be classified into six categories such as speaker identification, speaker verification, speaker detection, speaker segmentation, speaker clustering and speaker diarylation[55]. Speech signals contain speaker-specific information which can be extracted and fed to a machine learning algorithm to learn a specific pattern in it. MFCC and LPC are considered as the most two well-known features that have been widely used by researchers for speaker recognition applications, due to their capacity to capture the repetitive nature and efficiency of speech signals [55], [56]. Plenty of

research have claimed that the MFCC is an effective feature in correctly recognizing speakers. For instance, three features for automatic speech recognition were evaluated based on the Fuzzy approach including MFCC, dynamic time warping, and fast Fourier transform. Based on their result, the MFCC improves the performance of the fuzzy model compared to fast Fourier transformation features [57]. Al-Ali et al. enhanced forensic speaker verification based on the fusion features namely the MFCC and Discrete Wavelet Transform (DWT), where their models were evaluated in a noisy environment [58]. Abdul investigated the MFCC feature for speaker identification. He used the MFCC features to feed to the CNN and the result shows that the MFCC features could be used to train the CNN model to distinguish between speakers [7]. Despite the capacity of MFCC in capturing the characteristics of a speaker, the performance of MFCC degrades on complex speech datasets and in noise environments. For instance [59] present that the speaker recognition using MFCC and k-NN significantly degrades under noisy environment, and conclude that denoising the input signal can improve the result more when the highest MFCCs is adopted. To overcome this problem, in [60] a multi-channel training framework within the deep speaker embedding network was proposed for speaker recognition under reverberant and noisy environment. The method receives the time-, frequency-, and spatial-information from the multi-channel input to improve the robustness speaker embedding process. The work concludes that little increase in model parameters, can make the method significantly outperform i-vector with MFCC system with front-end signal enhancement. Additionally, Jahangir et al proposed fusion features based on MFCC and time-based features. The fusion features were fed to the DNN to identify speakers. The result shows that the limitation of the MFCC feature can be solved by this approach [56]. Table (4) displays the summarization of a further thirty papers that used the MFCC for speaker recognition.

3.1.2.2. Gender recognition over the phone call.

Gender recognition is the process of identifying a speaker's gender via analysis and comparison of patterns. Gender recognition is one of the effective techniques for biometric and has been a widely interested application for forensic teams. The most Common feature for gender recognition is the pitch feature of the speech. However, the literature, refers to the use of the MFCC features for gender recognition as well. For example, MFCC features were merged with the speaker's mean pitch for gender recognition. Both features were extracted from the acoustic signal and fed to the Gaussian Mixture Model[89]. The MFCC features were extracted from telephonic speech. The MFCC features were evaluated with several machine learning algorithms such as random forest, k-nearest neighbor, multilayer perceptron, naïve Bayes, and support vector machine [90]. Kang and Chang optimized the MFCC feature based on minimum classification error "SVM" and called weighted MFCC as a weight vector of the coefficients was optimized. The

weighted MFCC feature outperformed conventional MFCC [91]. More details are shown in Table (5).

3.1.3. Digital Forensic

Digital forensic can be defined as an application or a process which helps the forensic team to identify, analyze and extract important information regarding a specific crime or fakes from digital data. Usually, it is considered as robust evidence in court of law. Digital forensic applications vary depending on obtained information such as Network information and personal information about criminals.

3.1.3.1. Sentiment speech analysis

Sentiment speech detection is the analysis of the speech signal to extract subjective information in source material for identifying the mood of the speaker and customer that helps a business understand the social sentiment of their product and service during the monitoring of online chats. Generally, sentiment speech analysis classifies acoustic data into three statuses including positive, negative, and neutral categories [94]. The MFCC feature has been used a lot for sentiment speech analysis; for instance, Amiriparian et al. trained SVM based on the bag-of-audio-words (BoAW), which was obtained from the 39-dimensional MFCC feature. The 39-dimensional MFCCs consist of 12-MFCCs, 12-ΔMFCCs, 12-ΔΔMFCCs, Energy(E) , ΔE, and ΔΔE [95]. The 13, 65 and 130-Dimension MFCCs were also used to feed a sentiment model based on Gaussian Mixture Models (GMM) and Deep Neural Network (DNN) classifiers separately and the result shows that the 65 dimensions MFCC with the DNN outperformed the GMM model [96]. Authors in [97] studied Sentiment Analysis on Speaker Specific Speech by extracting the MFCC feature from the speech signal and feeding it to Dynamic Time Wrapping (DTW).

3.1.3.2. Fake speech detection

Although many recent technologies have been developed for voice recognition speaker verification, it still struggles and opens issues which is having a universal fake speech and it needs to be detected. The main aim of fake speech detection is to differentiate between a fake speech and a natural speech [98]. There are plenty of features that have been used for detecting fake speech such as MFCC and LPC. Sanchez et al. in [99] proposed a model based on the statistical classifier for synthetic speech and the MFCC was used as an authorized baseline. Lui et al. in [100] developed a model for fake Stereo Audio detection where the classifier was the SVM, and the feature was MFCC. Based on their result, the MFCC can detect stereo-faking audio very effectively. Authors in [101] developed a deep learning model. which was trained by the MFCC, ΔMFCC and ΔΔMFCC, for discriminating a synthesized speech from an actual human speech. they have found that both ΔMFCC and ΔΔMFCC can improve discrimination accuracy significantly compared to the MFCC performance alone. Table (6) summarizes the use MFCC in some other studies.

3.2. Medical applications

Nowadays, machine learning approaches have been involved in many healthcare applications and it helps to observe critical clues among huge data about the patient where the observation is beyond the scope of human capability, or it is difficult to know them in the shortest time [112]. Both traditional machine learning and deep learning approaches have been adopted by researchers to convert medical signals like ECG and EEG into clinical insight which is meaningful for the doctors. One of the crucial features is MFCC features, which was originally developed to extract vocal track features, however, shows the ability to extract useful features from medical signals to find out a clinical insight as it is mentioned below.

3.2.1. EEG analysis

An electroencephalogram (EEG) is a recording of the electronic action generated by the brain. EEG has a crucial role to identify or recognize brain activities and their means as well. Plenty of research is available that focuses on brain activities and there are still many challenges in this field. One of the vital challenges is obtaining a proper feature from the brain signal.

Based on the literature, some researchers proposed MFCC features for monitoring the brain activity and knowing the emotion of the patient or client, for example, Othman et al. extracted MFCC features from the EEG signal for identifying and recognizing the emotions. The MFCC feature then fed to the multi-layer perceptron to classify the emotions (happy, fear, sad and calm) based on the EEG signal and the result shows that the accuracy can gain up to 90%[113]. The performance of the EEG classification method was improved by combining 13 MFCCs with 19 articulatory features and then, the features were classified by gated recurrent unit [114]. Rajesh extracted 15, 25, and 35 MFCCs features from the EEG signal and fed them to an ANN model separately for classifying normal and abnormal activity. Based on his result, 25 MFCC components are fair enough to provide a maximum accuracy compared to 15 and 35 MFCCs [115]. Based on the information in the table (7), the MFCC is extracted from both EEG and acoustic source to diagnose brain activities.

3.2.2. ECG analysis

An electrocardiogram (ECG) signal is a recording of the electronic action generated by the heart. Doctors and health staff usually use ECG to check the status of the heart. An automatic ECG classification system is an experimental area and there are still many efforts and challenges in identifying and classifying several waveforms in the ECG signal [121]. Recently, the MFCC features have been widely used for this purpose; for example, Yusuf & Hidayat evaluated two well-known features which are 13 MFCCs and Discrete Wavelet transformation when fed to the kNN. The obtained result shows that the performance of the 13 MFCCs outperformed DWT [122]. Boussaa et al. claimed that the MFCC is a

robustness feature for ECG classification into normal and abnormal when it feeds into a multilayer perceptron [123]. Authors in [124] extracted two sets of features, which were MFCCs and Motifs, extracted from the ECG signal and the features were used to train the SVM individually. Based on their experiments, the best results were achieved when the MFCC trained the SVM. Table (8) shows the narration for some other research on this application.

3.2.3.   Disease detection application

Nowadays, the adoption of innovative technologies has been increasing in the field of medicine rapidly to detect signs of the illness in the earlier stage or to enhance and monitor the status of the patients. Behind any modern technologies in the medical application, there are several signal processing methods as they all work on the observation data. Recently, the MFCC has shown up in the detection of some diseases such as Parkinson's disease, voice disorder, vocal fold disorder, pathology detector, and Cough detection (see table (9)) which are described briefly below.

3.2.3.1. Parkinson disease:

Parkinson's disease is known as one of the brain diseases and it was described by James Parkinson in 1817. The brain cells, which produce and store dopamine, are subjected to a gradual loss and this is also called a neurological disorder. Parkinson's disease leads to some difficulties in the patient such as shaking, difficulty with walking, stiffness, and unbalance[129].

 Plenty of intelligent systems have been developed for classifying healthy and patient people by analyzing the voice as the disease hits the voice in the beginning and becomes impaired[130]. So, as it is related to the voice, MFCC features have been used a lot for this purpose, for example, 1-20 coefficients of the MFCC extracted from the voice of the vowel / a / and is fed to the Linear kernels SVM to detect the PD. The consequences of the experiment shows that the first 12 coefficients of the MFCC are the best classification accuracy achieved which was 91.17% [131]. A modification of the MFCC was done by changing the bandwidth of the Mel scaled bank filter to extract the region of interest [300 Hz, 1700 Hz] that are more related to the PD and then trained a radial basis function (RBF) network to classify healthily and patient [132]. Authors in [133] used Linear discriminant analysis to reduce the dimension of the MFCC coefficients and fed it to the SVM and the result shows that 10, 12 and 18 MFCC are more related to the PD signs.

3.2.3.2. Voice pathology disorders:

Voice pathology disorder is a problem on the vocal track voice, which appears in pitch, volume, tone and other qualities of a voice. The problem occurs whenever the vocal cords do not vibrate normally. Computer-aided tools have played a significant role in the diagnosis and identification of a lot of diseases including voice pathology disorders [134]. The MFCC as a feature, has been adopted to train

computer-aided tools for classifying normal voice and voice pathology disorders, for instance a model based on the MFCC feature was developed for the same purpose and the MFCC was extracted from the speech signal when the speech has undergone the distortion of an analog communications channel. The performance of the MFCC is still reasonable to be conducted. However, the performance of the model was degraded because of the added distortion [135]. Ali et al. used the MFCC with the Gaussian mixture model (GMM) to develop an automatic detection system that could differentiate between normal and pathological voices [136]. Enhanced the traditional MFCC and named adaptive weighted Thomson multi-taper MFCC. The enhanced MFCC was given to the GMM model as input for differentiating normal and disorder speech, and the performance of the obtained MFCC outperformed the traditional MFCC [137]. In 2018, the 19 MFCC were extracted from the speech signal and ANN was trained for discriminating between healthy and pathological voices, and the high classification rate achieved up to 99.96% [138].

3.2.3.3. Covid 19 detection via cough

Cough is a symptom of over thirty medical conditions including COVID-19. For more than three decades, the detection of medical conditions based on coughing has been interesting to researchers who work in this field [139][140]. The performance MFCC feature for classifying cough and ordinary speech was reported to be better than some other features such as Short-time Fourier transform (STFT) and Mel-scale filter banks (MFB) [141]. Recently, the MFCCs have been used as a feature for detecting whether the patient has COVID-19 or not, for example, the MFCC features were obtained from cough sounds and fed to the kNN to detect COVID-19 infected patients [142]. Bansal et al. trained a CNN model by the MFCCs, which was obtained from the cough sound, and classified into two classes such as Covid and Non-Covid classes [143].

3.3. Industry analysis

Reliability of machines or tools is gaining importance in the industry because of the need to decrease the possible loss of production whenever the machine experiences an abnormal situation during the working load. There is plenty of market assessment tools, which are utilized by analysts to understand the competitive dynamics of an industry, and they are used in industry [157]. A Health monitoring system is one of the well-known assessment tools that has been used to monitor a dynamic system and predict a failure in the early stage [158]. In condition monitoring systems whether it deals with vibration [159] or acoustic signal[160]. The MFCC is known as an effective feature to monitor tools or equipment and has been used quite recently for this purpose such as gear health monitoring, bearing health monitoring, turbine health monitoring, and Pump health monitoring. Table (10) provides more details on other industrial applications.

3.3.1. Gear health monitoring

A gear is a type of machine equipment that has teeth around cone-shaped or cylindrical surfaces. The teeth are designated around the shape with equal spacing. A force from one shaft, which is named the driving shaft, transmits to another shaft (driven shaft). Monitoring gears in rotation machinery devices have been focused on the industry as the failure in gear downs the entire system where the gears are involved in construction [161]. Various faults are related to the gears and many machine learning-based models have been developed to recognize these faults [162]. The MFCC has become one of the effective features in gear fault detection, for instance, Benkedjouh et al. extracted the MFCC feature and fed it to the SVM and claimed that the first three MFCC components contain the most defect information of gears [163]. However, based on the research of Abdul et al, 1-13 MFCC are more effective to be taken to train LSTM [164] and Jin et al. evaluated some sets of MFCCs (16, 21, 26, 31, 36, 41 MFCCs) by feeding them into a CNN model individually and the result shows that 41 MFCCs outperformed the others for gear fault detection. In conclusion, the dimension of the MFCC features is important for increasing the classification rate[165].

3.3.2. Bearing health monitoring

Bearings are the most widely used components in industrial rotating machinery, which consists of rolling elements, inner race, outer race, and separator. The bearing is used for rotating or linear shaft applications. There are plenty of types of bearing that are available in the industry, namely Ball Bearings (like Deep-Groove Ball Bearings and Angular Contact Ball Bearings) and Roller Bearings (like Cylindrical Roller Bearings and Tapered Roller Bearings). In earlier times, researchers used some statistical features to predict how the bearing such as mean, variance, standard deviation, root mean square, skewness and kurtosis is healthy?. However, these features are not sufficient to monitor all faults that might occur in bearings[166]

So many distinctive features have been extracted to overcome this problem such as MFCC feature and gamma tone spectrum coeffects. Nelwamondo and Marwala investigated the performance of a combination of features, which consisted of the MFCC and kurtosis and gave it to the Gaussian Mixture model. Based on their result, combining the Kurtosis can improve the performance of the MFCCs features by 5%[167]. Nelwamondo et al. used the first thirteen MFCCs feature as temporal feature and hidden Markov Models (HMM) was trained to classify bearing faults. they observed that the first 13 MFCCs is the optimal for bearings fault and beyond 13 does not increase the fault detection rate[166]. In another research, 12 MFCC coefficients were extracted from the vibration signal and then combined with both Wavelet Packet Decomposition Energy features and the Zero-crossing rate features. The mixture feature was fed to the support vector machine to classify gear faults [168]. Mingsi et al. however, studied two limitations of the MFCC to fault detection for rolling bearing

fault diagnosis. Firstly the Mel scale originally developed to speech signal based on human hearing and it is not suitable for the frequency distribution of rolling bearing vibration signal and the second, the ability of the MFCC in denoising. For these limitations, the adaptive frequency cepstrum coefficient (AFCC) method was proposed by [169] and fed to the XGboost algorithm for classification bearing faults. Table (10) displays some researches in which MFCC has been used as feature.

3.3.3. Turbine health monitoring

A turbine is a set of devices that harness the kinetic energy of some liquid such as water, air, steam, and turns this into the rotational motion of the device itself [173]. The turbine plays a vital role in converting power and many application tools have been developed for monitoring the turbine[174]. The essential requirement for any monitoring system based on machine learning is a dataset as decisions come out based on the data analysis. Based on the existing literature, MFCC has been used to obtain defected features in turbines, for example the MFCC and Code Excited Linear Prediction were used to extract a feature set from the acoustic signal. Then the feature of each technique was compared into different states which were fault and non-fault states[175]. Wang et al. optimized MFCC features, which were extracted from wind turbine sound signals, based on the entropy weight method for fault diagnosis of the wind turbine blade. The optimized feature was later clustered using the K-means clustering[176]. Kilic evaluated three sets of features, which are MFCC, spectrogram, and Mel spectrogram, by feeding them into three machine learning models such as LSTM, RNN and CNN. Based on Kilic's result the MFCC feature with CNN can get better than the other adopted model [177].

3.3.4. Pump health monitoring

A pump is a device which moves fluids from one place to the other by mechanical action and it is particularly important to monitor it well as any failure of pump will affect productivity and safety. There are many types of pumps which are available in the industrial field, and they are classified based work, namely dynamic, and positive displacement. MFCC as feature has been used to know the status of a pump for instance for fault isolation of Solenoid pumps, Akpudo and Hur employed a locally linear embedding (LLE) for reducing dimension of the MFCC features which obtained from vibration signal and the reduced MFCC feature fed to the SVM with Gaussian kernel[178]. In 2021, Akpudo and Hur proposed another model for Fault Detection and Isolation in Electromagnetic Pumps. They used the same feature, which was 12 MFCC (2nd to 13th MFCCs) features, with two other features namely first-order and the second-order differentials (12 Delta and 12 Delta-Delta features), which are driven from the MFCC feature. A feature selection technique known as rank-based recursive feature elimination, was then used to select the most effective features among 36-dimension and later given to the SVM classifier [170]. Table (11) shows some other studies that MFCC has been used in the industrial field.

4. SUMMARY

Table (12) shows a statistic obtained from the survey about the use of MFCC in various methods for different applications. Since the MFCC is mostly derived for audio-based applications, one can clearly observe that it has been applied frequently to acoustic signals. However, despite that, this cepstral feature has recently started to be used for non-acoustic signals in various areas such as medical and industrial applications. MFCC is reported to be a powerful feature hence it has mostly been adopted alone, much less been combined with other features. The reason of combination of MFCC with other features is mostly to overcome the sensitivity of MFCC in noise environment [109].

Most of the works in the literature followed the standard approach of using MFCC. For example, the coefficient in most of the works is adopted to be 13 or less, and third of the papers tried to enhance the MFCC computation parameters. However, the traditional way of MFCC computation followed the demand of audio signal modeling, which is not necessarily to be useful for non-audio signal such as ECG, EEG, and vibration. Even though the input signals of the adopted application in this paper are of time-series nature, most of researchers represented the MFCC in a global way by computing the statistics of the coefficients along with the signal frames. Despite the promising result of MFCC in the industrial field, table (12) shows that only around 13% of the papers that are using MFCC is applied to industrial application. This encourages more investigation of MFCC in this field.

The handcrafted features including the MFCC are mostly feeding the classical machine learning and rarely supply deep learning classifiers. The reason could be the nature of some deep learning methods that are able to extract deep features directly from the raw data.

If we take the reviewed papers as a sample of what has been conducted using the MFCC, we can observe the following:

- Most of the acoustic-based applications such as speech recognition, speaker recognition, gender recognition from speech and language recognition are using the MFCC alone more than the combination of it with other features as it is shown in the figure (4). However, the works that adopt the emotion recognition application are mostly using the combination of MFCC with other features. This could be an indication of the sparse nature of emotional information in the speech signal which may not be captured properly by the MFCC alone.

- Among the same works of acoustic application, the conducted works are mostly using the standard form of the MFCC, while around half of the works of the speaker recognition-based articles are adopting improving the MFCC parameters (see

figure (4)). This may be due to the diversity of the speakers' vocal tracts, which needs MFCC computation process modification.

- Although the MFCC is computed in a short-time feature, most of the works are using the global version of the feature representation rather than the time series one as illustrated in figure (5). The reason could be the issues that may appear when using the time series version of the feature representation such as the diversity of the input signal length. Additionally, the application itself can benefit from the time series representation more than other representations.

- Works that adopt the use of non-acoustic application are mostly using the standard form of MFCC computation. Since the standard parameters of the MFCC are derived for audio applications, optimizing the MFCC parameters may have significant impact on the recognition results (see figure (6)).

- Most of the publications are using Hamming window in the signal framing step. Very little are using Hanning window, and none is adopting the use of rectangle window, however, some are not mentioning the adopted window in their work (see figure (7)).

- Based on reviewing 187 articles, the MFCC has been used with various classifiers including suitable classifiers for the global representation of features (such as SVM, kNN, ANN and MLC) and classifiers that are used with time series representation (such as DTW, HMM and LSTM). The review shows that the most used classifier with the MFCC feature depends on the application. Table (13) presents the ratio of the most used classifier for each covered application in this review. It is clearly seen that SVM, and ANN are widely adopted with MFCC for their capability to find optimum separators among classes. SVM is a high-performance classifier with representative features and is not affected by the curse of dimensionality. On the other hand, ANN has the capability to learn the linear and non-linear characteristics of various classes from their data samples. Despite its importance, deep learning classifiers such as CNN have not been widely utilized with the MFCC, because CNN is capable of extracting learned features directly from the raw data. The nature of the application has its role to choose the classifier as well. For example, the ASR is a time series-based application, therefore, classifiers that detect the characteristics of the time-varying samples such as DTW, HMM and RNNs are more adopted for such an application. One can

conclude from these results that nothing special is
observed about the relation between the MFCC and
other classifiers since the most adopted classifiers
are those that are widely used with other
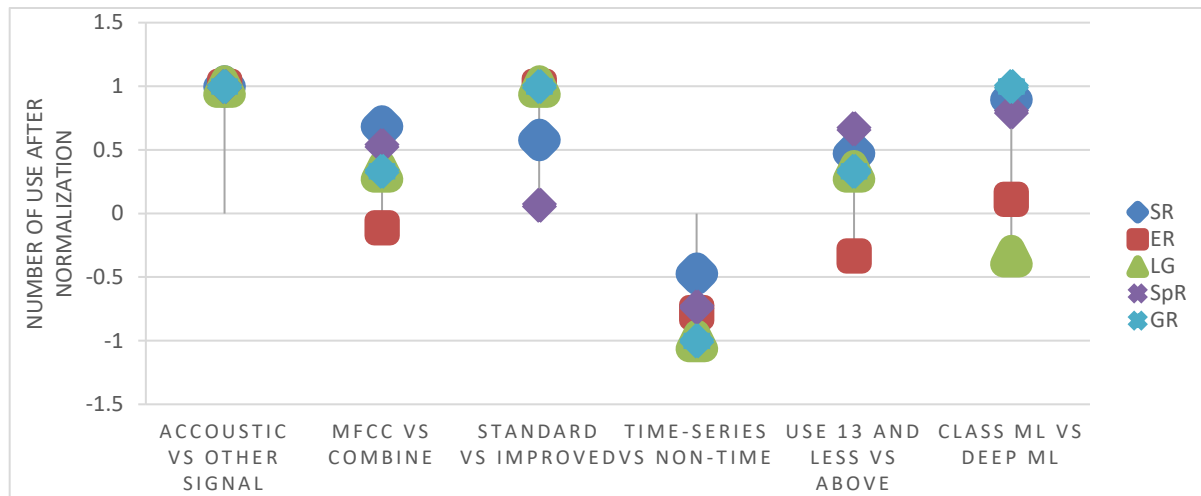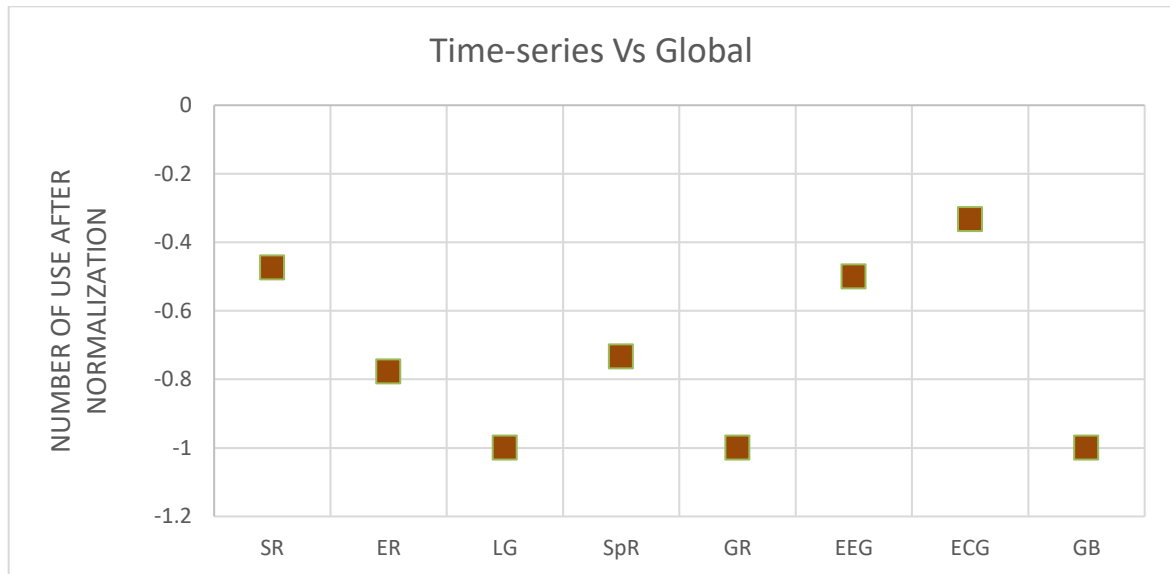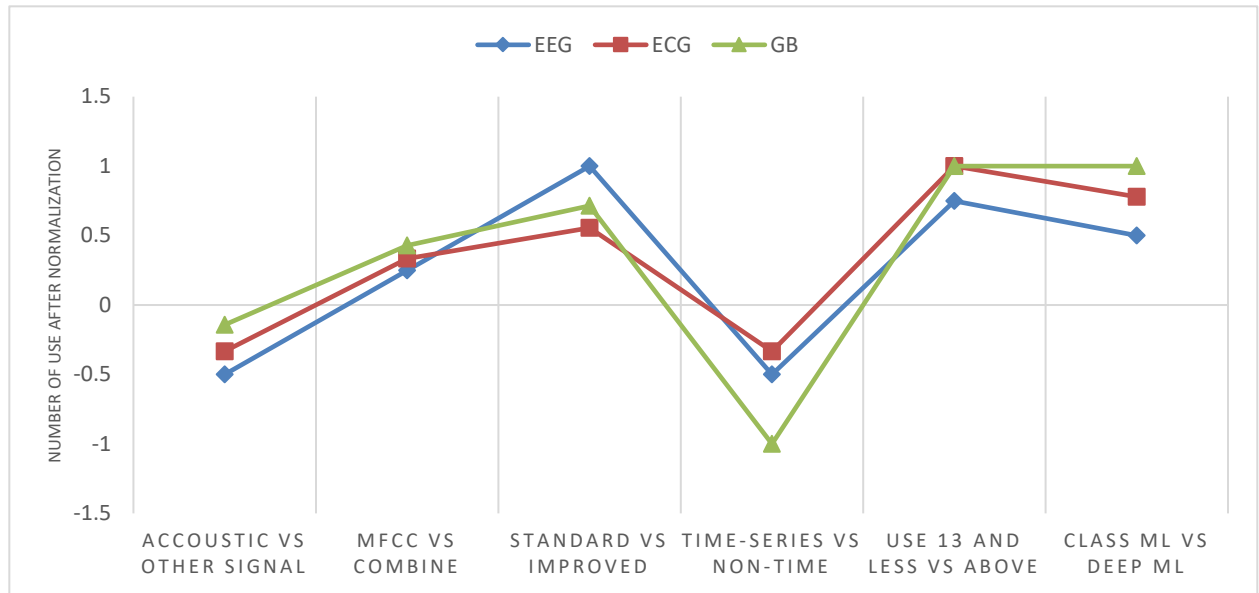applications as well.

*Figure 4 : The figure shows the number of papers of the first item in each pair (acoustic input vs other signals, MFCC
vs combined feature, standard MFCC vs improved version, time series vs global, using 13 MFCC vs more MFCC,
using classical machine learning vs using deep learning) normalized to one. (SR: Speech Recognition, ER: Emotion
Recognition, LG: Language Recognition, SpR: Speaker Recognition, and GR: Gender Recognition)*

Figure 5: The figure shows the number of papers of the first item in one pair (time series vs global) normalized to one. (SR: Speech Recognition, ER: Emotion Recognition, LG: Language Recognition, SpR: Speaker Recognition, GR: Gender Recognition and GB: Gearbox)

Figure 6: The figure shows the number of papers of the first item in each pair (acoustic input vs other signals, MFCC vs combined feature, standard MFCC vs improved version, time series vs global, using 13 MFCC vs more MFCC, using classical machine learning vs using deep learning) normalized to one.
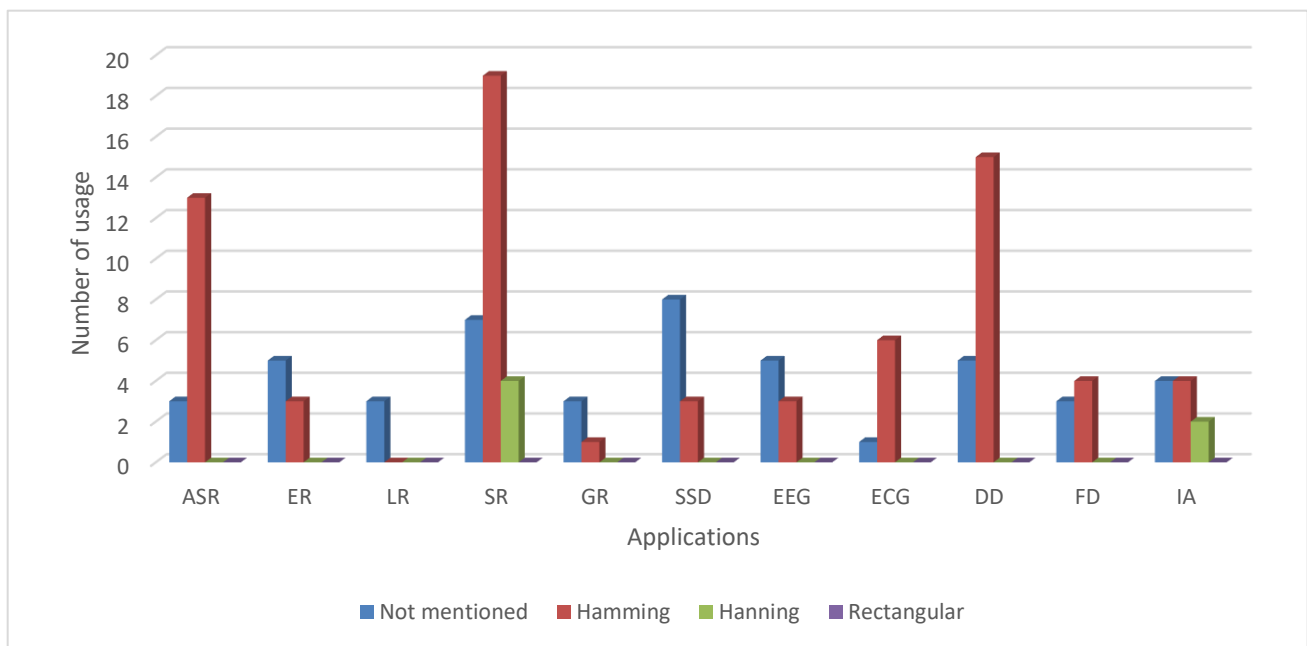
Figure 7: Utilized window types in various applications (ASR: Automatic Speech Recognition, ER: Emotion recognition, LR: Language recognition, SR: Speaker recognition, GR: Gender recognition, SSD: Sentiment speech detection, EEG: Electroencephalogram, ECG: Electrocardiogram, DD: Disease Detection, FD: Fault detection and IA: Industrial Application).

IEEE *Access*

Multidisciplinary : Rapid Review : Open Access Journal

5. CONCLUSION

This paper is a review of using MFCC in various applications and presents some issues regarding the computation of it. One can conclude from the review that covers 187 papers, that the MFCC is a widely used feature for acoustic-applications and a promising one for other applications such as EEG, ECG and industrial signals. However, there is no comprehensive investigation of MFCC for non-acoustics applications. Another finding is that most of the papers have not adapted the MFCC computation for non-acoustic applications. However, there are many indications that this modification may improve the performance of the non-acoustics models. Even though the MFCC is computed in a short time signal, however, most of the works has adopted the use of the global representation of the MFCC, which are mostly globalized by computing the statistics the features along the frames. The duality of time series versus non-time series is worth much more investigation to show their capability in modelling various applications.

REFERENCES

[1] L. J. P. Van Der Maaten, E. O. Postma, and H. J. Van Den Herik, "Dimensionality Reduction: A Comparative Review," J. Mach. Learn. Res., vol. 10, pp. 66–71, 2009, doi: 10.1080/13506280444000102.

[2] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," Proc. 2014 Sci. Inf. Conf. SAI 2014, no. August 2014, pp. 372–378, 2014, doi: 10.1109/SAI.2014.6918213.

[3] G. Krishna, C. Tran, M. Carnahan, Y. Han, and A. H. Tewfik, "Generating EEG features from acoustic features," Eur. Signal Process. Conf., vol. 2021-Janua, pp. 1100–1104, 2021, doi: 10.23919/Eusipco47968.2020.9287498.

[4] A. Sharma, M. Amarnath, and P. K. Kankar, "Feature extraction and fault severity classification in ball bearings," JVC/Journal Vib. Control, vol. 22, no. 1, pp. 176–192, 2016, doi: 10.1177/1077546314528021.

[5] C. Cremona, S. Yang, and L. Dieleman, "Overview of feature extraction techniques," IOMAC 2009 - 3rd Int. Oper. Modal Anal. Conf., no. September, pp. 381–390, 2009, doi: 10.2139/ssrn.3568122.

[6] B. Liang, S. D. Iwnicki, and Y. Zhao, "Application of power spectrum, cepstrum, higher order spectrum and neural network analyses for induction motor fault diagnosis," Mech. Syst. Signal Process., vol. 39, no. 1–2, pp. 342–360, 2013, doi: 10.1016/j.ymssp.2013.02.016.

[7] Abdul, Zrar Khalid. "Kurdish speaker identification based on one dimensional convolutional neural network." Computational Methods for Differential Equations 7.4 (Special Issue) (2019): 566-572.

[8] X. Zhao and D. Wang, "Analyzing noise robustness of MFCC and GFCC features in speaker identification," ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., pp. 7204–7208, 2013, doi: 10.1109/ICASSP.2013.6639061.

[9] Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvet, D., Fissore, L., Laface, P., Mertins, A., Ris, C. and Rose, R.., "Automatic speech recognition and speech variability: A review," Speech Commun., vol. 49, no. 10–11, pp. 763–786,

2007, doi: 10.1016/j.specom.2007.02.006.

[10] K. S. Rao and K. E. Manjunath, Speech recognition using articulatory and excitation source features. Springer, 2017.

[11] I. K. Timotius, Computational Methods for Gait Analysis in Rodents. FAU University Press, 2020.

[12] H. A. Feldman, N. Kaiser, and J. A. Peacock, "Power-spectrum analysis of three-dimensional redshift surveys," Astrophys. J., vol. 426, p. 23, doi 1994, doi: 10.1086/174036.

[13] H. Yin, V. Hohmann, and C. Nadeu, "Acoustic features for speech recognition based on Gammatone filterbank and instantaneous frequency," Speech Commun., vol. 53, no. 5, pp. 707–715, 2011, doi: 10.1016/j.specom.2010.04.008.

[14] G. Strang, "The discrete cosine transform," SIAM Rev., vol. 41, no. 1, pp. 135–147, 1999, doi: 10.1137/S0036144598336745.

[15] Heimann, T., Van Ginneken, B., Styner, M.A., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G. and Bello, F., "Comparison and evaluation of methods for liver segmentation from CT datasets," IEEE Trans. Med. Imaging, vol. 28, no. 8, pp. 1251–1265, 2009, doi: 10.1109/TMI.2009.2013851.

[16] S. Ramakrishnan, "Recognition of Emotion from Speech: A Review," in Speech Enhancement, Modeling and Recognition-Algorithms and Applications, 2012.

[17] A. Revathi, C. Ravichandran, P. Saisiddarth, and G. S. R. Prasad, "Isolated Command Recognition Using MFCC and Clustering Algorithm," SN Comput. Sci., vol. 1, no. 2, pp. 1–7, 2020, doi: 10.1007/s42979-020-0093-x.

[18] S. K. Gaikwad, B. W. Gawali, and P. Yannawar, "A Review on Speech Recognition Technique," Int. J. Comput. Appl., vol. 10, no. 3, pp. 16–24, 2010, doi: 10.5120/1462-1976.

[19] S. V.Chapaneri, "Spoken Digits Recognition using Weighted MFCC and Improved Features for Dynamic Time Warping," Int. J. Comput. Appl., vol. 40, no. 3, pp. 6–12, 2012, doi: 10.5120/5022-7167.

[20] S. S. and T. Y. Chadawan Ittichaichareon, "Speech recognition using MFCC," in International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012), 2012, pp. 135–138.

[21] S. Dhingra, G. Nijhawan, and P. Pandit, "Isolated speech recognition using MFCC and DTW," Int. J. Adv. …, vol. 2, no. 8, pp. 4085–4092, 2013, [Online]. Available: http://ijareeie.com/upload/2013/august/20P_ISOLATED.pdf.

[22] S. M. Omer, J. A. Qadir, and Z. K. Abdul, "Uttered Kurdish digit recognition system," J. Univ. Raparin, vol. 6, no. 2, pp. 78–85, 2019, doi: 10.26750/vol(6).no(2).paper5.

[23] E. S. Wahyuni, "Arabic speech recognition using MFCC feature extraction and ANN classification," Proc. - 2017 2nd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2017, vol. 2018-Janua, pp. 22–25, 2018, doi: 10.1109/ICITISEE.2017.8285499.

[24] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, "Speech Recognition Using Deep Neural Networks: A Systematic Review," IEEE Access, vol. 7, pp. 19143–19165, 2019, doi: 10.1109/ACCESS.2019.2896880.

[25] B. Zhen, X. Wu, Z. Liu, and H. Chi, "On the importance of components of the MFCC in speech and speaker recognition," Beijing Daxue Xuebao Ziran Kexue Ban/Acta Sci. Nat. Univ. Pekin., vol. 37, no. 3, pp. 371–378, 2001.

[26] A. S. Haq, M. Nasrun, C. Setianingsih, and M. A. Murti, "Speech recognition implementation using MFCC and DTW algorithm for home automation," Int. Conf. Electr. Eng. Comput. Sci. Informatics, vol. 7, no. October, pp. 78–85, 2020, doi: 10.11591/eecsi.v7.2041.

[27] K. R. A. Zade, C. Ardil, and S. S. Rustamov, "Investigation of Combined use of MFCC and LPC Features in Speech Recognition Systems," World Acad. Sci. Eng. Technol., vol. 3, no. 1, pp. 74–80, 2006.

[28] N. Dave, "Feature Extraction Methods LPC , PLP and MFCC In Speech Recognition," Int. J. Adv. Res. Eng. Technol., vol. 1, no. Vi, pp. 1–5, 2013.

[29] V. Z. Këpuska and H. A. Elharati, "Robust Speech Recognition System Using Conventional and Hybrid Features of MFCC, LPCC, PLP, RASTA-PLP and Hidden Markov Model Classifier in Noisy Conditions," J. Comput. Commun., vol. 03, no. 06, pp. 1–9, 2015, doi: 10.4236/jcc.2015.36001.

[30] H. M. Soe Naing, R. Hidayat, R. Hartanto, and Y. Miyanaga, "Discrete wavelet denoising into mfcc for noise suppressive in automatic speech recognition system," Int. J. Intell. Eng. Syst., vol. 13, no. 2, pp. 74–82, 2020, doi: 10.22266/ijies2020.0430.08.

[31] B. Wu, S. Sakti, S. Nakamura, and R. Aip, "MFCC-DPGMM Features for Enhancing Low-Resource ASR ∗," vol. 2, pp. 2–3, 2019.

[32] S. D. and P. Roy, Emerging Technology in Modelling and Graphics, vol. 937. 2020.

[33] B. Paul, S. Bera, R. Paul, and S. Phadikar, "Bengali Spoken Numerals Recognition by MFCC and GMM Technique," in Advances in Electronics, Communication and Computing, Springer, 2021, pp. 85–96.

[34] M. D. Skowronski and J. G. Harris, "Increased MFCC filter bandwidth for noise-robust phoneme recognition," ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., vol. 1, no. September, 2002, doi: 10.1109/icassp.2002.5743839.

[35] F. Zheng, G. Zhang, and Z. Song, "Comparison of different implementations of MFCC," J. Comput. Sci. Technol., vol. 16, no. 6, pp. 582–589, 2001, doi: 10.1007/BF02943243.

[36] S. Umesh and R. Sinha, "A study of filter bank smoothing in MFCC features for recognition of children's speech," IEEE Trans. Audio, Speech Lang. Process., vol. 15, no. 8, pp. 2418–2430, 2007, doi: 10.1109/TASL.2007.906194.

[37] D. R. KS, M. D. Rudresh, and G. Shashibhushsan, "Comparative Performance Analysis for Speech Digit Recognition based on MFCC and Vector Quantization," Glob. Transitions Proc., 2021.

[38] N. J.Nalini and S. Palanivel, "Emotion Recognition in Music Signal using AANN and SVM," Int. J. Comput. Appl., vol. 77, no. 2, pp. 7–14, 2013, doi: 10.5120/13364-0958.

[39] H. Aouani and Y. Ben Ayed, "Emotion recognition in speech using MFCC with SVM, DSVM and auto-encoder," in 2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 2018, pp. 1–5.

[40] A. Al-talabani, "Automatic Speech Emotion Recognition-Feature space Dimensionality and Classification Challenges," Dr. thesis, no. January, 2015.

[41] A. B. Kandali, A. Routray, and T. K. Basu, "Emotion recognition from Assamese speeches using MFCC features and GMM classifier," IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON, 2008, doi: 10.1109/TENCON.2008.4766487.

[42] H. Ibrahim, C. K. Loo, and F. Alnajjar, "Speech Emotion Recognition by Late Fusion for Bidirectional Reservoir Computing with Random Projection," IEEE Access, vol. 9, pp. 122855–122871, 2021, doi: 10.1109/ACCESS.2021.3107858.

[43] P. P. Dahake, K. Shaw, and P. Malathi, "Speaker dependent speech emotion recognition using MFCC and Support Vector Machine," Int. Conf. Autom. Control Dyn. Optim. Tech. ICACDOT 2016, pp. 1080–1084, 2017, doi: 10.1109/ICACDOT.2016.7877753.

[44] H. Aouani and Y. Ben Ayed, "Speech Emotion Recognition with deep learning," Procedia Comput. Sci., vol. 176, pp. 251–260, 2020, doi: 10.1016/j.procs.2020.08.027.

[45] K. V. K. Kishore and P. K. Satish, "Emotion Recognition in Speech Using MFCC and Wavelet Features," in 3rd IEEE International Advance Computing Conference, 2012, pp. 842–847.

[46] A. Milton, S. Sharmy Roy, and S. Tamil Selvi, "SVM Scheme for Speech Emotion Recognition using MFCC Feature," Int. J. Comput. Appl., vol. 69, no. 9, pp. 34–39, 2013, doi: 10.5120/11872-7667.

[47] N. J. Nalini and S. Palanivel, "Music emotion recognition: The combined evidence of MFCC and residual phase," Egypt. Informatics J., vol. 17, no. 1, pp. 1–10, 2016, doi: 10.1016/j.eij.2015.05.004.

[48] S. G. Koolagudi, D. Rastogi, and K. S. Rao, "Identification of language using mel-frequency cepstral coefficients (MFCC)," Procedia Eng., vol. 38, pp. 3391–3398, 2012, doi: 10.1016/j.proeng.2012.06.392.

[49] A. M. Warohma, P. Kurniasari, S. Dwijayanti, B. Y. Suprapto, and others, "Identification of regional dialects using mel frequency cepstral coefficients (MFCCs) and neural network," in 2018 International Seminar on Application for Technology of Information and Communication, 2018, pp. 522–527.

[50] B. Tawaqal and S. Suyanto, "Recognizing Five Major Dialects in Indonesia Based on MFCC and DRNN," J. Phys. Conf. Ser., vol. 1844, no. 1, 2021, doi: 10.1088/1742-6596/1844/1/012003.

[51] E. Mansour, M. S. Sayed, A. M. Moselhy, and A. A. Abdelnaiem, "LPC and MFCC Performance Evaluation with Artificial Neural Network for Spoken Language Identification," Int. J. Signal Process. Image Process. Pattern Recognit., vol. 6, no. 3, pp. 55–66, 2013.

[52] A. Dogra, A. Kaul, and R. N. Sharma, "Automatic Recognition of Dialects of Himachal Pradesh Using MFCC GMM," Proc. IEEE Int. Conf. Signal Process. Control, vol. 2019-Octob, pp. 134–137, 2019, doi: 10.1109/ISPCC48220.2019.8988336.

[53] A. K. Al-talabani, Z. K. Abdul, and A. A. Ameen, "Kurdish Dialects and Neighbor Languages Automatic Recognition," no. 1, pp. 20–23, 2017, doi: 10.14500/aro.10167.

[54] S. Shon, A. Ali, and J. Glass, "Convolutional Neural Network and Language Embeddings for End-to-End Dialect Recognition," pp. 98–104, 2018, doi: 10.21437/odyssey.2018-14.

[55] R. Mohd Hanifa, K. Isa, and S. Mohamad, "A review on speaker recognition: Technology and challenges," Comput. Electr. Eng., vol. 90, no. January, p. 107005, 2021, doi: 10.1016/j.compeleceng.2021.107005.

[56] R. Jahangir et al., "Text-Independent Speaker Identification through Feature Fusion and Deep Neural Network," IEEE Access, vol. 8, pp. 32187–32202, 2020, doi: 10.1109/ACCESS.2020.2973541.

[57] V. Singh and N. Meena, "Engine Fault Diagnosis using DTW, MFCC and FFT," Proc. First Int. Conf. Intell. Hum. Comput. Interact., pp. 83–94, 2009, doi: 10.1007/978-81-8489-203-1_6.

[58] A. K. H. Al-Ali, B. Senadji, and G. R. Naik, "Enhanced forensic speaker verification using multi-run ICA in the presence of environmental noise and reverberation conditions," Proc. 2017 IEEE Int. Conf. Signal Image Process. Appl. ICSIPA 2017, vol. 5, pp. 174–179, 2017, doi: 10.1109/ICSIPA.2017.8120601.

[59] F. K. Faek and A. K. Al-Talabani, "Speaker recognition from noisy spoken sentences," Int. J. Comput. Appl., vol. 70, no. 20, 2013.

[60] D. Cai, X. Qin, and M. Li, "Multi-Channel Training for End-to-End Speaker Recognition Under Reverberant and Noisy Environment.," in Interspeech, 2019, pp. 4365–4369.

[61] S. Nakagawa, K. Asakawa, and L. Wang, "Speaker Recognition by Combining MFCC and Phase Information," INTERSPEECH 2007, pp. 27–31, 2007.

[62] D. Hosseinzadeh and S. Krishnan, "Combining vocal source and MFCC features for enhanced speaker recognition performance using GMMs," 2007 IEEE 9Th Int. Work. Multimed. Signal Process. MMSP 2007 - Proc., pp. 365–368, 2007, doi: 10.1109/MMSP.2007.4412892.

[63] S. Chakroborty and G. Saha, "Improved text-independent Speaker Identification using fused MFCC & IMFCC feature sets based on Gaussian filter," World Acad. Sci. Eng. Technol., vol. 35, no. 1, pp. 613–621, 2009, doi: 10.5281/zenodo.1073555.

[64] M. A. Hossan, S. Memon, and M. A. Gregory, "A novel approach for MFCC feature extraction," 4th Int. Conf. Signal Process. Commun. Syst. ICSPCS'2010 - Proc., no. July 2014, 2010, doi: 10.1109/ICSPCS.2010.5709752.

[65] M. Sahidullah and G. Saha, "Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition," Speech Commun., vol. 54, no. 4, pp. 543–565, 2012, doi: 10.1016/j.specom.2011.11.004.

[66] C. Kumar, F. Rehman, S. Kumar, A. Mehmood, and G. Shabir, "Analysis of MFCC and BFCC in a Speaker Identification System," no. March, 2018, doi: 10.1109/ICOMET.2018.8346330.

[67] K. Sri Rama Murty and B. Yegnanarayana, "Combining evidence from residual phase and MFCC features for speaker recognition," IEEE Signal Process. Lett., vol. 13, no. 1, pp. 52–55, 2006, doi: 10.1109/LSP.2005.860538.

[68] K. S. Ahmad, A. S. Thosar, J. H. Nirmal, and V. S. Pande, "A unique approach in text independent speaker recognition using MFCC feature sets and probabilistic neural network," ICAPR 2015 - 2015 8th Int. Conf. Adv. Pattern Recognit., 2015, doi: 10.1109/ICAPR.2015.7050669.

[69] P. B. 2 Kailashnath J K 1, "kannada speaker recognition using mel frequency cepstral coefficients (mfcc) and artificial neural network," Int. J. Creat. Res. Thoughts, vol. 6, no. 1, p. 1411, 2018.

[70] B. B. V. Rao and H. G. Chandrakanth, "Secured speech industrial automation based on Raspberry Pi and IoT," Int. J. Intell. Eng. Syst., vol. 11, no. 6, pp. 234–243, 2018, doi: 10.22266/IJIES2018.1231.23.

[71] N. P. Patel and A. Kale, "Optimize Approach to Voice Recognition Using IoT," in 2018 International Conference On Advances in Communication and Computing Technology, ICACCT 2018, 2018, pp. 251–256, doi: 10.1109/ICACCT.2018.8529622.

[72] T. Ganchev, N. Fakotakis, and G. Kokkinakis, "Comparative Evaluation of Various MFCC Implementations on the Speaker Verification Task," October, vol. 1, no. 3, pp. 191–194, 2005, doi: 10.1.1.75.8303.

[73] C. H. Lee, C. H. Chou, C. C. Han, and R. Z. Huang, "Automatic recognition of animal vocalizations using averaged MFCC and linear discriminant analysis," Pattern Recognit. Lett., vol. 27, no. 2, pp. 93–101, 2006, doi: 10.1016/j.patrec.2005.07.004.

[74] M. Pleva, E. Kiktova, J. Juhar, and P. Bours, "Acoustical user identification based on MFCC analysis of keystrokes," Adv. Electr. Electron. Eng., vol. 13, no. 4, pp. 309–313, 2015, doi: 10.15598/aeee.v13i4.1466.

[75] H. S. Bae, H. J. Lee, and S. G. Lee, "Voice recognition-based on adaptive MFCC and deep learning for embedded systems," J. Inst. Control. Robot. Syst., vol. 22, no. 10, pp. 797–802, 2016, doi: 10.5302/J.ICROS.2016.16.0136.

[76] Z. Wu and Z. Cao, "Improved MFCC-based feature for robust speaker identification," Tsinghua Sci. Technol., vol. 10, no. 2, pp. 158–161, 2005, doi: 10.1016/S1007-0214(05)70048-1.

[77] Fu, Qinjian, Danjue Lv, Yan Zhang, Haifeng Xu, Wang Yao, Jiali Zi, and Jiang Liu., "Research on Crane Sound Clustering of MFCC Based on HHT," J. Phys. Conf. Ser., vol. 1693, no. 1, 2020, doi: 10.1088/1742-6596/1693/1/012134.

[78] A. Chowdhury and A. Ross, "Fusing MFCC and LPC Features Using 1D Triplet CNN for Speaker Recognition in Severely Degraded Audio Signals," IEEE Trans. Inf. Forensics Secur., vol. 15, no. c, pp. 1616–1629, 2020, doi: 10.1109/TIFS.2019.2941773.

[79] A. Ashar, M. S. Bhatti, and U. Mushtaq, "Speaker Identification Using a Hybrid CNN-MFCC Approach," 2020 Int. Conf. Emerg. Trends Smart Technol. ICETST 2020, 2020, doi: 10.1109/ICETST49965.2020.9080730.

[80] S. Chen and Y. Luo, "Speaker Verification Using MFCC and Support Vector Machine," Lect. Notes Eng. Comput. Sci., vol. 2174, no. 1, pp. 532–535, 2009.

[81] M. Sahidullah and G. Saha, "A novel windowing technique for efficient computation of MFCC for speaker recognition," IEEE Signal Process. Lett., vol. 20, no. 2, pp. 149–152, 2013, doi: 10.1109/LSP.2012.2235067.

[82] M. O. F. Engineering, "Speaker Identification of Disguised Voices Using Mfcc Statistical Moment and Svm," THAPAR Univ., 2017.

[83] Y. Alsouda, S. Pllana, and A. Kurti, "IoT-based urban noise identification using machine learning: Performance of SVM, KNN, bagging, and random forest," ACM Int. Conf. Proceeding Ser., vol. Part F1481, pp. 62–67, 2019, doi: 10.1145/3312614.3312631.

[84] J. Martinez, H. Perez, E. Escamilla, and M. M. Suzuki, "Speaker recognition using Mel Frequency Cepstral Coefficients (MFCC) and Vector quantization (VQ) techniques," CONIELECOMP 2012 - 22nd Int. Conf. Electron. Commun. Comput., no. February, pp. 248–251, 2012, doi: 10.1109/CONIELECOMP.2012.6189918.

[85] V. Tiwari, "MFCC and its applications in speaker recognition," Int. J. Emerg. Technol., vol. 1, no. 1, pp. 19–22, 2010.

[86] P. S. Singh and V. Khanna, "A MFCC based Novel approach of User Authentication in IOT," in 2nd International Conference on Emerging Trends in Engineering & Applied Science (ICETEAS' 19), 2020, no. 1, pp. 28–33.

[87] M. J. Alam, T. Kinnunen, P. Kenny, P. Ouellet, and D. O'Shaughnessy, "Multitaper MFCC and PLP features for speaker verification using i-vectors," Speech Commun., vol. 55, no. 2, pp. 237–251, 2013, doi: 10.1016/j.specom.2012.08.007.

[88] L. Muda, M. Begam, and I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques," J. Comput., vol. 2, no. 3, pp. 138–143, 2010, [Online]. Available: http://arxiv.org/abs/1003.4083.

[89] H. Ting, Y. Yingchun, and W. Zhaohui, "Combining MFCC and pitch to enhance the performance of the gender recognition," Int. Conf. Signal Process. Proceedings, ICSP, vol. 1, pp. 3–6, 2006, doi: 10.1109/ICOSP.2006.345541.

[90] J. Ahmad, M. Fiaz, S. Kwon, M. Sodanil, B. Vo, and S. W. Baik, "Gender Identification using MFCC for Telephone Applications - A Comparative Study," vol. 3, no. 5, 2016, [Online]. Available: http://arxiv.org/abs/1601.01577.

[91] S. I. Kang and J. H. Chang, "Discriminative weight training-based optimally weighted MFCC for gender identification," IEICE Electron. Express, vol. 6, no. 19, pp. 1374–1379, 2009, doi: 10.1587/elex.6.1374.

[92] E. Fokoue, Z. Ma, and E. Fokoué, "Speaker Gender Recognition via MFCCs and SVMs," 2013. [Online]. Available: http://scholarworks.rit.edu/article/1749.

[93] F. Faek, "Objective Gender and Age Recognition from Speech Sentences," Aro, Sci. J. Koya Univ., vol. 3, no. 2, pp. 24–29, 2015, doi: 10.14500/aro.10072.

[94] T. Kumar, M. Mahrishi, and S. Nawaz, "A Review of Speech Sentiment Analysis Using Machine Learning," in Proceedings of Trends in Electronics and Health Informatics, 2022, pp. 21–28.

[95] S. Amiriparian, N. Cummins, S. Ottl, M. Gerczuk, and B. Schuller, "Sentiment analysis using image-based deep spectrum features," 2017 7th Int. Conf. Affect. Comput. Intell. Interact. Work. Demos, ACIIW 2017, vol. 2018-Janua, pp. 26–29, 2018, doi: 10.1109/ACIIW.2017.8272618.

[96] S. H. V. B, V. Mohan, and K. Urkalan, "Multimodal Sentiment Analysis Using Deep Neural Networks Harika," vol. 30, no. April, pp. 13–23, 2017, doi: 10.1007/978-3-319-58130-9.

[97] S. Maghilnan and M. R. Kumar, "Sentiment analysis on speaker specific speech data," Proc. 2017 Int. Conf. Intell. Comput. Control. I2C2 2017, vol. 2018-Janua, pp. 1–5, 2018, doi: 10.1109/I2C2.2017.8321795.

[98] S. Camacho, D. M. Ballesteros, and D. Renza, "Fake Speech Recognition Using Deep Learning," in Workshop on Engineering Applications, 2021, pp. 38–48.

[99] J. Sanchez, I. Saratxaga, I. Hernaez, E. Navas, and D. Erro, "A cross-vocoder study of speaker independent synthetic speech detection using phase information," Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH, no. September, pp. 1663–1667, 2014, doi: 10.21437/interspeech.2014-393.

[100] T. Liu, D. Yan, R. Wang, N. Yan, and G. Chen, "Identification of fake stereo audio using svm and cnn," Inf., vol. 12, no. 7, 2021, doi: 10.3390/info12070263.

[101] A. K. Singh, P. Singh, and K. Nathwani, "Using Deep Learning Techniques and Inferential Speech Statistics for AI Synthesised Speech Recognition," pp. 1–13, 2021, [Online]. Available: http://arxiv.org/abs/2107.11412.

[102] A. Martin, D. Charlet, and L. Mauuary, "Robust speech/non-speech detection using LDA applied to MFCC," ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., vol. 1, pp. 237–240, 2001, doi: 10.1109/ICASSP.2001.940811.

[103] O. Chia Ai, M. Hariharan, S. Yaacob, and L. Sin Chee, "Classification of speech dysfluencies with MFCC and LPCC features," Expert Syst. Appl., vol. 39, no. 2, pp. 2157–2165, 2012, doi: 10.1016/j.eswa.2011.07.065.

[104] T. Kinnunen and E. Chernenko, "Voice activity detection using MFCC features and support vector machine," Int. Conf. Speech, vol. 2, no. 2, pp. 555–561, 2007, [Online]. Available: http://www.cs.joensuu.fi/pages/tkinnu/webpage/pdf/SVMVAD_SPECOM2007.pdf.

[105] Y. Alsouda, S. Pllana, and A. Kurti, "A Machine Learning Driven IoT Solution for Noise Classification in Smart Cities," 2018, [Online]. Available: http://arxiv.org/abs/1809.00238.

[106] G. Muhammad and K. Alghathbar, "Environment recognition for digital audio forensics using MPEG-7 andmel cepstral features," J. Electr. Eng., vol. 62, no. 4, pp. 199–205, 2011, doi: 10.2478/v10187-011-0032-0.

[107] A. Abordo and J. Liao, "VOICE COMMAND RECOGNITION SYSTEM BASED ON MFCC AND DTW," Electr. Eng., no. December 2010, p. 107, 2003.

[108] J. Ebbers, A. Nelus, R. Martin, and R. Häb-Umbach, "Evaluation of Modulation-MFCC Features and DNN Classification for Acoustic Event Detection," Dtsch. Jahrestagung fur Akust., 2017.

[109] S. Jin, X. Wang, L. Du, and D. He, "Evaluation and modeling of automotive transmission whine noise quality based on MFCC and CNN," Appl. Acoust., vol. 172, p. 107562, 2021, doi: 10.1016/j.apacoust.2020.107562.

[110] Fu, Qinjian, Danju Lv, Yan Zhang, Haifeng Xu, Wang Yao, Jiali Zi, and Jiang Liu, "Research on Ethnic Original Ecological Music Clustering of MFCC Based on HHT," J. Phys. Conf. Ser., vol. 1848, no. 1, 2021, doi: 10.1088/1742-6596/1848/1/012035.

[111] C. J. Tralie, "Early MFCC and HPCP fusion for robust cover song identification," Proc. 18th Int. Soc. Music Inf. Retr. Conf. ISMIR 2017, pp. 294–301, 2017.

[112] I. Konenko, "Machine learning for medical diagnosis: History, state of the art and perspective," Artif. Intell. Med., vol. 23, no. 1, pp. 89–109, 2001, [Online]. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed5&NEWS=N&AN=2001260608.

[113] M. Othman, A. Wahab, and R. Khosrowabadi, "MFCC FOR ROBUST EMOTION DETECTION USING EEG," in IEEE 9th Malaysia International Conference on Communications, 2009, no. December, p. 5532.

[114] G. Krishna, C. Tran, M. Carnahan, Y. Han, and A. H. Tewfik, "Improving EEG based continuous speech recognition," arXiv, pp. 1–5, 2019.

[115] G. Rajesh, "Analysis of Mfcc Features for Eeg Signal Classification," Int. J. Adv. Signal Image Sci., vol. 2, no. 2, p. 14, 2016, doi: 10.29284/ijasis.2.2.2016.14-20.

[116] and R. R.-V. Cesar Ortiz-Echeverri, Omar Paredes, J. Sebastian Salazar-Colores, Juvenal Rodr´ıguez-Res´endiz, "A Comparative Study of Time and Frequency Features for EEG Classificatio," in IFMBE Proceedings, 2020, vol. 75, pp. 92–97, doi: 10.1007/978-3-030-30648-9_13.

[117] P. P. Mini, T. Thomas, and R. Gopikakumari, "EEG based direct speech BCI system using a fusion of SMRT and MFCC/LPCC features with ANN classifier," Biomed. Signal Process. Control,

vol. 68, p. 102625, 2021, doi:
https://doi.org/10.1016/j.bspc.2021.102625.

[118] G. Krishna, C. Tran, M. Carnahan, and A. H. Tewfik, "Advancing speech synthesis using EEG," Int. IEEE/EMBS Conf. Neural Eng. NER, vol. 2021-May, pp. 199–204, 2021, doi: 10.1109/NER49283.2021.9441306.

[119] B. Farzam and J. Shirazi, "The Diagnosis of Heart Diseases Based on PCG Signals using MFCC Coefficients and SVM Classifier," Int. J. Innov. Sciennce, Eng. Technol., vol. 1, no. 10, pp. 654–659, 2014.

[120] T. G. Kwon and J. H. Shin, "An analysis of Brain waves characteristic with Acupuncture Feedback using MFCC," 10th ACIS Conf. Softw. Eng. Artif. Intell. Netw. Parallel/Distributed Comput. SNPD 2009, conjunction with IWEA 2009 WEACR 2009, pp. 385–390, 2009, doi: 10.1109/SNPD.2009.81.

[121] E. J. da S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: A survey," Comput. Methods Programs Biomed., vol. 127, pp. 144–164, 2016, doi: 10.1016/j.cmpb.2015.12.008.

[122] S. A. A. Yusuf and R. Hidayat, "MFCC feature extraction and KNN classification in ECG signals," 2019 6th Int. Conf. Inf. Technol. Comput. Electr. Eng. ICITACEE 2019, pp. 1–5, 2019, doi: 10.1109/ICITACEE.2019.8904285.

[123] M. Boussaa, I. Atouf, M. Atibi, and A. Bennis, "ECG signals classification using MFCC coefficients and ANN classifier," Proc. 2016 Int. Conf. Electr. Inf. Technol. ICEIT 2016, no. May, pp. 480–484, 2016, doi: 10.1109/EITech.2016.7519646.

[124] D. M. Nogueira, C. A. Ferreira, E. F. Gomes, and A. M. Jorge, "Classifying Heart Sounds Using Images of Motifs, MFCC and Temporal Features," J. Med. Syst., vol. 43, no. 6, 2019, doi: 10.1007/s10916-019-1286-5.

[125] H. Hussain, Hadri and Ting, Chee-Ming and Numan, Fuad and Ibrahim, M. Nasir and Izan, Nur Fariza and Mohammad, M M and Sh-Hussain, "Analysis of ECG Biosignal Recognition for Client Identifiction," Encycl. Commun. theory, pp. 12–14, doi: 10.1109/ICSIPA.2017.8120572.

[126] S. Chauhan, P. Wang, C. Sing Lim, and V. Anantharaman, "A computer-aided MFCC-based HMM system for automatic auscultation," Comput. Biol. Med., vol. 38, no. 2, pp. 221–233, 2008, doi: 10.1016/j.compbiomed.2007.10.006.

[127] S. Zokaee and K. Faez, "Human Identification Based on ECG and Palmprint," Int. J. Electr. Comput. Eng., vol. 2, no. 2, 2012, doi: 10.11591/ijece.v2i2.292.

[128] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, and H. Fan, "Heart sound classification based on improved MFCC features and convolutional recurrent neural networks," Neural Networks, vol. 130, pp. 22–32, 2020, doi: 10.1016/j.neunet.2020.06.015.

[129] P. Shrivastava, A. Shukla, P. Vepakomma, N. Bhansali, and K. Verma, "A survey of nature-inspired algorithms for feature selection to identify Parkinson's disease," Comput. Methods Programs Biomed., vol. 139, pp. 171–179, 2017, doi: 10.1016/j.cmpb.2016.07.029.

[130] S. Hegde, S. Shetty, S. Rai, and T. Dodderi, "A Survey on Machine Learning Approaches for Automatic Detection of Voice Disorders," J. Voice, vol. 33, no. 6, pp. 947.e11-947.e33, 2019, doi: 10.1016/j.jvoice.2018.07.014.

[131] A. Benba, A. Jilbab, A. Hammouch, and S. Sandabad, "Voiceprints analysis using MFCC and SVM for detecting patients with Parkinson's disease," Proc. 2015 Int. Conf. Electr. Inf. Technol. ICEIT 2015, pp. 300–304, 2015, doi:

10.1109/EITech.2015.7163000.

[132] S. S. Upadhya, A. N. Cheeran, and J. H. Nirmal, "Discriminating Parkinson diseased and healthy people using modified MFCC filter bank approach," Int. J. Speech Technol., vol. 22, no. 4, pp. 1021–1029, 2019, doi: 10.1007/s10772-019-09647-0.

[133] A. Rahman, S. S. Rizvi, A. Khan, A. A. Abbasi, S. U. Khan, and T. S. Chung, "Parkinson's disease diagnosis in cepstral domain using MFCC and dimensionality reduction with SVM classifier," Mob. Inf. Syst., vol. 2021, 2021, doi: 10.1155/2021/8822069.

[134] F. T. Al-Dhief et al., "Voice Pathology Detection Using Machine Learning Technique," in 2020 IEEE 5th International Symposium on Telecommunication Technologies, ISTT 2020 - Proceedings, 2020, no. December, pp. 99–104, doi: 10.1109/ISTT50966.2020.9279346.

[135] J. I. Godino-llorente and C. Fredouille, "MFCC-based remote pathology detection on speech transmitted through the telephone channel - Impact of Linear Distortions: Band Limitation, Frequency Response and Noise," pp. 41–48, 2011, doi: 10.5220/0001534200410048.

[136] Z. Ali, M. Alsulaiman, G. Muhammad, I. Elamvazuthi, and T. A. Mesallam, "Vocal fold disorder detection based on continuous speech by using MFCC and GMM," in 2013 7th IEEE GCC Conference and Exhibition, GCC 2013, 2013, no. April 2014, pp. 292–297, doi: 10.1109/IEEEGCC.2013.6705792.

[137] Ö. Eskidere and A. Gürhanli, "Voice Disorder Classification Based on Multitaper Mel Frequency Cepstral Coefficients Features," Comput. Math. Methods Med., vol. 2015, 2015, doi: 10.1155/2015/956249.

[138] Smitha, S. Shetty, S. Hegde, and T. Dodderi, "Classification of Healthy and Pathological voices using MFCC and ANN," in Proceedings of 2018 2nd International Conference on Advances in Electronics, Computers and Communications, ICAECC 2018, 2018, pp. 1–5, doi: 10.1109/ICAECC.2018.8479441.

[139] M. Cohen-Mcfarlane, R. Goubran, and F. Knoefel, "Novel Coronavirus Cough Database: NoCoCoDa," IEEE Access, vol. 8, pp. 154087–154094, 2020, doi: 10.1109/ACCESS.2020.3018028.

[140] B. W. Schuller et al., "The INTERSPEECH 2021 Computational Paralinguistics Challenge: COVID-19 Cough, COVID-19 Speech, Escalation & Primates," pp. 1–5, 2021, doi: 10.21437/interspeech.2021-19.

[141] I. D. S. Miranda, A. H. DIacon, and T. R. Niesler, "A Comparative Study of Features for Acoustic Cough Detection Using Deep Architectures∗," Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, pp. 2601–2605, 2019, doi: 10.1109/EMBC.2019.8856412.

[142] M. Melek, "Diagnosis of COVID-19 and non-COVID-19 patients by classifying only a single cough sound," Neural Comput. Appl., pp. 0–2, 2021, doi: 10.1007/s00521-021-06346-3.

[143] V. Bansal, G. Pahwa, and N. Kannan, "Cough classification for COVID-19 based on audio mfcc features using convolutional neural networks," 2020 IEEE Int. Conf. Comput. Power Commun. Technol. GUCON 2020, pp. 604–608, 2020, doi: 10.1109/GUCON48875.2020.9231094.

[144] L. Boquete, D. E. Bocanegra, and L. De-Santiago, "Recognition of Brain Structures from MER-Signals Using Dynamic MFCC Analysis and a HMC Classifie," IFMBE Proc. 41 XIII Mediterr. Conf. Med. Biol. Eng. Comput. 2013, pp. 325–329, 2014, doi: 10.1007/978-3-319-00846-2.

[145] L. Liu, B. Li, R. Zhao, W. Yao, M. Shen, and J. Yang, "A novel method for broiler abnormal sound detection using WMFCC and HMM," J. Sensors, vol. 2020, 2020, doi: 10.1155/2020/2985478.

[146] A. Zabidi, L. Y. Khuan, W. Mansor, I. M. Yassin, and R. Sahak, "Optimization of MFCC parameters using Particle Swarm Optimization for diagnosis of infant hypothyroidism using Multi-Layer Perceptron," 2010 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC'10, pp. 1417–1420, 2010, doi: 10.1109/IEMBS.2010.5626712.

[147] S. S. Upadhya, A. N. Cheeran, and J. H. Nirmal, "Thomson Multitaper MFCC and PLP voice features for early detection of Parkinson disease," Biomed. Signal Process. Control, vol. 46, pp. 293–301, 2018, doi: 10.1016/j.bspc.2018.07.019.

[148] A. Benba, A. Jilbab, and A. Hammouch, "Analysis of multiple types of voice recordings in cepstral domain using MFCC for discriminating between patients with Parkinson's disease and healthy people," Int. J. Speech Technol., vol. 19, no. 3, pp. 449–456, 2016, doi: 10.1007/s10772-016-9338-4.

[149] Z. Soumaya, B. D. Taoufiq, B. Nsiri, and A. Abdelkrim, "Diagnosis of Parkinson disease using the wavelet transform and MFCC and SVM classifier," in Proceedings of 2019 IEEE World Conference on Complex Systems, WCCS 2019, 2019, vol. 4, pp. 1–6, doi: 10.1109/ICoCS.2019.8930802.

[150] V. Srinivasan, V. Ramalingam, and P. Arulmozhi, "Artificial Neural Network Based Pathological Voice Classification Using Mfcc Features," Int. J. Sci. Environ. Technol., vol. 3, no. 1, pp. 291–302, 2014.

[151] T. Grzywalski et al., "Parameterization of Sequence of MFCCs for DNN-based voice disorder detection," Proc. - 2018 IEEE Int. Conf. Big Data, Big Data 2018, pp. 5247–5251, 2019, doi: 10.1109/BigData.2018.8622012.

[152] E. Rejaibi, A. Komaty, F. Meriaudeau, S. Agrebi, and A. Othmani, "MFCC-based Recurrent Neural Network for Automatic Clinical Depression Recognition and Assessment from Speech," Signal Process. Image Commun., 2019, [Online]. Available: http://arxiv.org/abs/1909.07208.

[153] P. Kadambi et al., "towards a wearable cough detector based on neural networks," pp. 2161–2165, 2018.

[154] S. Gupta, R. S. Shukla, R. K. Shukla, and R. Verma, "Deep learning bidirectional LSTM based detection of prolongation and repetition in stuttered speech using weighted MFCC," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 9, pp. 345–356, 2020, doi: 10.14569/IJACSA.2020.0110941.

[155] S. Y. Jung, C. H. Liao, Y. S. Wu, S. M. Yuan, and C. T. Sun, "Efficiently classifying lung sounds through depthwise separable cnn models with fused stft and mfcc features," Diagnostics, vol. 11, no. 4, 2021, doi: 10.3390/DIAGNOSTICS11040732.

[156] M. B. Alsabek, I. Shahin, and A. Hassan, "Studying the Similarity of COVID-19 Sounds based on Correlation Analysis of MFCC," Proc. 2020 IEEE Int. Conf. Commun. Comput. Cybersecurity, Informatics, CCCI 2020, 2020, doi: 10.1109/CCCI49893.2020.9256700.

[157] I. S. Candanedo, E. H. Nieves, S. R. González, M. T. S. Martín, and A. G. Briones, "Machine learning predictive model for industry 4.0," Commun. Comput. Inf. Sci., vol. 877, no. July, pp. 501–510, 2018, doi: 10.1007/978-3-319-95204-8_42.

[158] P. Henriquez, J. B. Alonso, M. A. Ferrer, and C. M. Travieso, "Review of automatic fault diagnosis systems using audio and vibration signals," IEEE Trans. Syst. Man, Cybern. Syst., vol. 44, no. 5, pp. 642–652, 2014, doi: 10.1109/TSMCC.2013.2257752.

[159] T. Wang, Q. Han, F. Chu, and Z. Feng, "Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox : A review," Mech. Syst. Signal Process., vol. 126, pp. 662–685, 2019, doi: 10.1016/j.ymssp.2019.02.051.

[160] Y. Qu, E. Bechhoefer, D. He, and J. Zhu, "A new acoustic emission sensor based gear fault detection approach," Int. J. Progn. Heal. Manag., vol. 4, no. SPECIAL ISSUE 2, pp. 1–14, 2013.

[161] J. R. Colbourne, The geometry of involute gears. Springer Science & Business Media, 2012.

[162] G. Dalpiaz, A. Rivola, and R. Rubini, "Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears," Mech. Syst. Signal Process., vol. 14, no. 3, pp. 387–412, 2000, doi: 10.1006/mssp.1999.1294.

[163] A. M. Benkedjouh T., Chettibi T., Saadouni Y., Gearbox Fault Diagnosis Based on Mel-Frequency Cepstral Coefficients and Support Vector Machine. Springer International Publishing, 2018.

[164] Z. K. Abdul, A. K. Al-Talabani, and D. O. Ramadan, "A hybrid temporal feature for Gear Fault Diagnosis using the Long Short Term Memory," IEEE Sens. J., vol. 23, no. 23, pp. 14444–114452, 2020, doi: 10.1109/jsen.2020.3007262.

[165] S. Jin, X. Wang, L. Du, and D. He, "Evaluation and modeling of automotive transmission whine noise quality based on MFCC and CNN," Appl. Acoust., vol. 172, p. 107562, 2021, doi: https://doi.org/10.1016/j.apacoust.2020.107562.

[166] AlShorman, O., Irfan, M., Saad, N., Zhen, D., Haider, N., Glowacz, A., "A Review of Artificial Intelligence Methods for Condition Monitoring and Fault Diagnosis of Rolling Element Bearings for Induction Motor," Shock Vib., vol. 2020, no. Cm, 2020, doi: 10.1155/2020/8843759.

[167] F. V. Nelwamondo and T. Marwala, "Faults detection using Gaussian mixture models, mel-frequency cepstral coefficients and kurtosis," Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern., vol. 1, no. 2, pp. 290–295, 2006, doi: 10.1109/ICSMC.2006.384397.

[168] J. Zhang, Q. Jiang, and F. Chang, "Fault diagnosis method based on MFCC Fusion and SVM," 2018 IEEE Int. Conf. Inf. Autom. ICIA 2018, no. August, pp. 1617–1622, 2018, doi: 10.1109/ICInfA.2018.8812417.

[169] M. Qi, R. Zhou, Q. Zhang, and Y. Yang, "Feature Classification Method of Frequency Cepstrum Coefficient Based on Weighted Extreme Gradient Boosting," IEEE Access, vol. 9, pp. 72691–72701, 2021, doi: 10.1109/ACCESS.2021.3079286.

[170] U. E. Akpudo and J. W. Hur, "A cost-efficient mfcc-based fault detection and isolation technology for electromagnetic pumps," Electron., vol. 10, no. 4, pp. 1–21, 2021, doi: 10.3390/electronics10040439.

[171] J. Fang and P. Liu, "Personalized Power Consumption Control Method for Intelligent Buildings That Perceiving Speech Recognition Identity," IOP Conf. Ser. Earth Environ. Sci., vol. 510, no. 2, 2020, doi: 10.1088/1755-1315/510/2/022007.

[172] A. K. Kemalkar and V. K. Bairagi, "Engine fault diagnosis using sound analysis," Int. Conf. Autom. Control Dyn. Optim. Tech. ICACDOT 2016, pp. 943–946, 2017, doi: 10.1109/ICACDOT.2016.7877726.

[173] M. P. Boyce, Gas turbine engineering handbook. Elsevier, 2011.

[174] W. Qiao and D. Lu, "A Survey on Wind Turbine Condition Monitoring and Fault Diagnosis - Part II: Signals and Signal

Processing Methods," IEEE Trans. Ind. Electron., vol. 62, no. 10, pp. 6546–6557, 2015, doi: 10.1109/TIE.2015.2422394.

[175] C. C. Kuo et al., "mfcc and celpto detect turbine engine faults," vol. 2, no. 12, 2014.

[176] Y. Wang, J. Tan, X. Qin, Z. Hu, Y. Jing, and G. Han, "Windturbine blade fault diagnosis based on MFCC feature optimization strategy," IOP Conf. Ser. Earth Environ. Sci., vol. 675, no. 1, 2021, doi: 10.1088/1755-1315/675/1/012074.

[177] M. Kilic, "Fault Detection of Wind Turbines Using Deep Learning," no. May, 2021.

[178] U. E. Akpudo and J. W. Hur, "Intelligent Solenoid Pump Fault Detection based on MFCC Features, LLE and SVM," 2020 Int. Conf. Artif. Intell. Inf. Commun. ICAIIC 2020, no. February, pp. 404–408, 2020, doi: 10.1109/ICAIIC48513.2020.9065282.

[179] G. Zhai, J. Chen, C. Li, and G. Wang, "Pattern recognition approach to identify loose particle material based on modified MFCC and HMMs," Neurocomputing, vol. 155, pp. 135–145, 2015, doi: 10.1016/j.neucom.2014.12.039.

[180] I. Ahmad and K. H. Chang, "Analysis on MIMO transmit diversity and multiplexing techniques for ship ad-hoc networks under a maritime channel model in coastline areas," Int. Conf. Inf. Commun. Technol. Converg. ICT Converg. Technol. Lead. Fourth Ind. Revolution, ICTC 2017, vol. 2017-Decem, no. October, pp. 945–947, 2017, doi: 10.1109/ICTC.2017.8190820.

[181] J. Lee, H. Choi, D. Park, Y. Chung, H. Y. Kim, and S. Yoon, "Fault detection and diagnosis of railway point machines by sound analysis," Sensors (Switzerland), vol. 16, no. 4, 2016, doi: 10.3390/s16040549.

[182] Z. K. Abdul, A. K. Al-Talabani, and D. O. Ramadan, "Highly Accurate Gear Fault Diagnosis based on Support Vector Machine," Journal of Vibration Engineering & Technologies, in-press.

[183] F. Lyu, Z. Yang, L. Wang, Q. Zhou, S. Hou, and X. Li, "Behavior Anomaly Detection Fused with Features of Mel Frequency Cepstrum Coefficients," in 2020 International Conference on Sensing, Measurement Data Analytics in the era of Artificial Intelligence (ICSMD), 2020, pp. 89–93, doi: 10.1109/ICSMD50554.2020.9261684.

[184] C. Honggang, X. Mingyue, F. Chenzhao, S. Renjie, and L. Zhe, "Mechanical Fault Diagnosis of GIS Based on MFCCs of Sound Signals," in 2020 5th Asia Conference on Power and Electrical Engineering (ACPEE), 2020, pp. 1487–1491, doi: 10.1109/ACPEE48638.2020.9136284.

[185] X. Liu, D. Pei, G. Lodewijks, Z. Zhao, and J. Mei, "Acoustic signal based fault detection on belt conveyor idlers using machine learning," Adv. Powder Technol., vol. 31, no. 7, pp. 2689–2698, 2020, doi: 10.1016/j.apt.2020.04.034.

[186] J. de A. A. Marques and H. M. G. Castelo Branco, "Localização de Faltas em Linha de Transmissão Utilizando Coeficientes de Frequência Mel Cesptrais (MFCC) e Redes Neurais Artificiais (RNA)," 2020, doi: 10.48011/asba.v2i1.1037.

[187] T. Tsuji, K. Sato, and S. Sakaino, "Contact Feature Recognition Based on MFCC of Force Signals," IEEE Robot. Autom. Lett., vol. 6, no. 3, pp. 5153–5158, 2021, doi: 10.1109/LRA.2021.3072035.

Zrar Kh. Abdul was born in Kirkuk, Iraq, in May, 1984. He received his Bsc degree in 2009 from Koya University, Erbil/Iraq and M. degree in 2012 from the University of Sheffield, Sheffield/ Unite Kingdom. He is currently pursuing his Ph.D. degree in the Erbil Technical Engineering College at Erbil Polytechnic University, Erbil/ Iraq. His research interests include deep learning and its applications in condition-based maintenance for rotating machinery. ID: https://orcid.org/0000-0001-6586-8642

Abdulbasit K. Al-Talabani was born in Kirkuk city, Iraq in 1977. He received the B.S. in Mathematics from Salahaddin University, M.S. and PhD in computer science from Koya University, Kurdistan Region, Iraq in 2006 and the University of Buckingham, the UK in 2016 respectively. From 2003 to 2006 he was Research assistance at Education College, Koya University. Since 2006 he has been an assistant lecturer, then a lecturer at the Software Engineering Department, Koya University. His research interest includes machine learning, speech analysis, deep learning and vision

ID: https://orcid.org/0000-0001-6328-204X