Received April 21, 2021, accepted May 4, 2021, date of publication May 21, 2021, date of current version June 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3082565

Voice Pathology Detection and Classification by Adopting Online Sequential Extreme Learning Machine

FAHAD TAHA AL-DHIEF[1], (Graduate Student Member, IEEE), MARINA MAT BAKI[2], NURUL MU'AZZAH ABDUL LATIFF[1], (Senior Member, IEEE), NIK NOORDINI NIK ABD. MALIK[1], (Member, IEEE), NASEER SABRI SALIM[3], MUSATAFA ABBAS ABBOOD ALBADER[4], NOR MUZLIFAH MAHYUDDIN[5], (Member, IEEE), AND MAZIN ABED MOHAMMED[6]

[1]Faculty of Engineering, School of Electrical Engineering, Universiti Teknologi Malaysia (UTM), Johor Bahru 81310, Malaysia
[2]Department of Otorhinolaryngology, Faculty of Medicine, Universiti Kebangsaan Malaysia Medical Centre, Kuala Lumpur 56000, Malaysia
[3]Computing and Information Technology, Sohar University, Sohar 311, Oman
[4]CAIT, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia
[5]School of Electrical and Electronic Engineering, Universiti Sains Malaysia, Nibong Tebal 14300, Malaysia
[6]College of Computer Science and Information Technology, University of Anbar, Ramadi 45654, Iraq

Corresponding author: Nurul Mu'azzah Abdul Latiff (nurulmuazzah@utm.my)

This work was supported in part by the Ministry of Higher Education, Malaysia, in part by the Universiti Teknologi Malaysia (UTM) under Grant R.J130000.2651.18J74, in part by the Universiti Kebangsaan Malaysia (UKM) under Grant GP-2020-K014873, in part by the Research Code under Project FF-2020-262, and in part by the Faculty of Engineering, Universiti Teknologi Malaysia (UTM).

ABSTRACT In the last decade, the implementation of machine learning algorithms in the analysis of voice disorder is paramount in order to provide a non-invasive voice pathology detection by only using audio signal. In spite of that, most recent systems of voice pathology work on a limited acoustic database. In other words, the systems use one vowel, such as /a/, and ignore sentences and other vowels when analyzing the audio signal. Other key issues that should be considered in the systems are accuracy and time consumption of an algorithm. Online Sequential Extreme Learning Machine (OSELM) is one of the machine learning algorithms that can be regarded as a rapid and accurate algorithm in the classification process. Therefore, this paper presents a voice pathology detection and classification system by using OSELM algorithm as a classifier, and Mel-frequency cepstral coefficient (MFCC) as a featured extraction. In this work, the voice samples were taken from the Saarbrücken voice database (SVD). This system involves two parts of the database; the first part includes all voices in SVD with sentences and vowels /a/, /i/, and /u/, which are uttered in high, low, and normal pitches; and the second part utilizes voice samples of the common three types of pathologies (cyst, polyp, and paralysis) based on the vowel /a/ that is produced in normal pitch. The experimental results have shown that OSELM was able to achieve the highest accuracy up to 91.17%, 94% of precision, and 91% of recall. Furthermore, OSELM obtained 87%, 87.55%, and 97.67% for f-measure, G-mean, and specificity, respectively. The proposed system also presents a high ability to achieve detection and classification results in real-time clinical applications.

INDEX TERMS Machine learning, healthcare, voice pathology detection, pathologies classification, OSELM, MFCC, SVD.

I. INTRODUCTION

Voice pathology analysis is considered a very significant field in the healthcare area. Many people are suffering from voice troubles due to several reasons, such as extreme damage of certain organs, air pollution, smoking, and stress [1]. In a recent study, it has been found that more than 7.5 million people in the United States are suffering from voice pathology [2]. Moreover, approximately 25% of the world population suffer from voice problems as their careers have forced them to speak extremely louder than normal, for example, teachers, singers, auctioneers, lawyers, and actors, who need

The associate editor coordinating the review of this manuscript and approving it for publication was Wenbing Zhao.

to work arduously on the voice [3]. Voice pathology surveillance systems have obtained a huge interest because of the increasing risks of pathological voice problems [4]. Audio evaluation of voice pathologies is an imperative tool for clinicians where it is considered a useful assisting tool for medical doctors in terms of identifying voice disorders, especially when these voice problems are identified at early stages [5]. The evaluation of voice pathology systems can be performed by three different methods which are objective, subjective, and perceptive. The objective evaluation does not need any particular tools. When the algorithm is proper, the results are always impartial [6]. Objective evaluation can be used for initial screening only, and the final decision should come from medical doctors. For subjective evaluation, it requires particular tools and trained doctors, hence, it incurs higher cost [7]. Subjective evaluation varies from doctor to doctor where it depends on a doctor's expertise [8]. The perceptive evaluation involves identification of voice disorder by a group of trained professionals [9]. These professionals listen to a patient's speech and then evaluate the speech in order to determine whether the patient has a voice disorder or normal using the GRBAS scale [10]. However, perceptive evaluation suffers from the reliance on listener's experience and different knowledge of the judges. Consequently, the study of speech signal processing of pathological voices by objective evaluation becomes an important topic for researchers as it aims to reduce medical laboratory work in diagnosing pathological speeches. Furthermore, it provides a non-invasive method of diagnosis which is more comfortable for patients, faster, as well as cost-effective [11].

Machine learning algorithms can serve as objective evaluation tools for speech processing in order to detect pathological voices from acoustic recordings. In fact, it could be very valuable to detect voice disorders at an early stage or to measure the quality of voice before and after a surgery. Machine learning algorithms present techniques, methods, and tools that can be used as assistants for solving diagnosis problems in many medical specialties [12], [13]. Some of the different machine learning algorithms that have been utilized in speech analysis of voice pathology systems are Extreme Learning Machine (ELM) [14], Support Vector Machine (SVM) [15], and Gaussian Mixture Model (GMM) [16]. Thus, machine learning algorithms have proven their effectiveness and efficiency in differentiating pathological voices from normal voices. However, some of these algorithms still suffer low classification accuracy, long-time execution, or workload in the voice pathology monitoring systems. In addition, most machine-learning techniques require retraining the whole dataset when there is new data to be tested, and therefore, consume more time. This issue is considered a major current issue since it causes higher delay in order to obtain results. Another concerning issue is that most studies in voice pathology systems are based on small database, deal with limited vowels, such as /a/, and disregard other databases, such as the vowels /i/ and /u/, and sentences. The limitations of voice pathology detection systems can be summarized as follows:

- The majority of studies focus only on the detection of voice pathology and ignore pathological classification tasks.
- Most systems work on one voice data only, such as the vowel /a/, where other vowels and sentences are ignored.
- The number of voice database for healthy and pathological samples is limited.
- Machine learning classifiers still suffer from low accuracy rate.
- Systems of voice pathology are evaluated in terms of accuracy, specificity, and sensitivity only.

Hence, it is crucial to develop a reliable voice pathology detection system based on machine learning that is able to manage all these issues. The contributions of this paper are as follows:

- The OSELM algorithm is proposed as a classifier in the detection and classification of voice pathologies.
- The proposed system makes use of healthy and pathological voice samples from SVD which considers sentences and vowels /a/, /i/, and /u/ that are produced in three different pitches.
- The proposed system uses an expansive number of healthy and pathological voice samples to train and test the OSELM classifier.
- The system aims to detect and classify three common voice pathologies which are cyst, polyp, and paralysis from the voice samples.
- Several evaluation measurements are used to evaluate and show the efficiency of the proposed method.
- To the best of our knowledge, no research has used OSELM algorithm in voice pathology classification, and this study is the first to utilize OSELM classifier in voice pathology detection and classification using sentences and different vowels.

This paper is organized as follows; Section II discusses the drawbacks of the state-of-the-art in the systems of voice pathology detection and classification; Section III describes the proposed methods in terms of SVD, MFCC, and OSELM classifier; Section IV presents the experimental results of OSELM for all SVD in general and also for three common pathologies in particular. Finally, Section V concludes the paper.

II. RELATED WORK

Recent studies have addressed many voice pathology detection models and their main objective was to obtain improved results and higher accuracy in terms of the classification process. These proposed systems have examined voice quality metrics such as shimmer, jitter, noise harmonic ratio, signal-to-noise ratio, and the glottal-to-noise ratio [17], [18]. There are also other methods that study different acoustic features, for instance, Mel-Frequency Cepstral Coefficient (MFCC) [19], Perceptual Linear Prediction (PLP) [20], and Multidimensional Voice Program (MDVP) [21]. These acoustic features are extracted from speech signals to be

processed and analyzed. Along with these features, there are well-known classifiers used to classify voice signals such as GMM, Hidden Markov Models (HMM) [22], Artificial Neural Network (ANN) [23], and SVM. To perform the objective evaluation of voice pathology, several databases were used by the researchers such as the Arabic Voice Pathology Database (AVPD) [24], Saarbrücken Voice Database (SVD) [25], and Massachusetts Eye and Ear Infirmary Database (MEEI) [26]. These databases are widely used in the voice pathology field and they have many normal and pathological voice samples. Nonetheless, the studies in the detection of voice pathology models are still at the early stage and insufficient. Therefore, it is imperative to investigate other machine learning algorithms for the implementation of the voice pathology area, taking into account a wider range of voice samples and databases.

A new system is proposed for automatic voice pathology detection [27]. In this system, Kullback–Leibler Divergence (KLD) is applied to the frame's histogram (H-KLD), and its spectrum is modified, known as Higher Amplitude Suppression Spectrum (HASS-KLD). H-KLD and HASS are used for measuring the distribution of speech signal frames in order to use a low number of parameters and obtain high accuracy. In particular, the H-KLD is used to measure the difference between two probability distributions of the speech signal frames. At the same time, HASS-KLD is used to capture dynamic aspects of the speech signals which include the short-term voice spectrum. Then, H-KLD and HASS-KLD are used for feature extraction to be fed to two Generalized Extreme Value (GEV) classifiers. In addition, MFCC is also used for feature extraction and the results are fed to GMM classifier. These three classifiers are combined and fed to Gaussian Naive Bayes (GNB) classifier that is responsible for the classification of the voice signal. Furthermore, MEEI database is adopted in this method with 53 and 173 samples for normal and pathological voices, respectively. This method is able to achieve up to 99.55% of accuracy for voice pathology assessment. However, this method is time-consuming since it employs three feature extraction techniques and four classifiers.

Another work has exploited three different feature extractions which are Signal Entropy (SH), Signal Energy (SE), and Zero-Crossing Rate (ZCR) with Discriminative Paraconsistent Machine (DPM) classifier for voice pathology system [28]. These features are combined to form a feature vector of voice signal and they are fed to DPM classifier. For the voice samples, SVD is used in this method. The samples are divided into four Classes (Cs), where C1 includes 10 voices affected by Reinke edema, C2 consists of 10 patients suffering from laryngitis, C3 contains 10 voices for both laryngitis and Reinke edema, and C4 has 10 normal voices. The maximum accuracy of this method can reach up to 95% for the classification process. Nonetheless, only a small size of voice samples is trained and tested using DPM classifier. Assessments and detections of the glottal signals for voice disorders are presented in [29]. This method aims to extract the parameters of glottal signals by applying the inverse filtering technique. Aparat Software is used to obtain glottal signal parameters where these parameters are extracted into time domain and frequency domain. Furthermore, k-nearest neighboring (k-NN) and SVM are used to classify the voice signal. The pathological and healthy voice samples are taken from SVD. The accuracy of SVM is 98.5%, while the accuracy of k-NN is 88.2% as reported in this paper. However, this method only utilizes a small set of voice samples.

Deep learning model is applied in the system of automatic voice pathology detection by using smart healthcare framework using a mobile platform [30]. Smartphones are used to record voice signals of users before they are uploaded into a cloud server. The voice signals are processed, analyzed and classified into three parallel models using Convolutional Neural Network (CNN) in the cloud. Upon completion of the classification process, the results are sent to stakeholders such as hospitals and doctors. The work in [30] is based on SVD with 1342 pathological samples and 686 healthy samples. It is shown in this paper that parallel CNNs are able to achieve up to 95.5% of accuracy. However, the classification is limited to sustained vowel /a/ that is uttered in normal pitch only.

A new system is proposed in [31] to analyze and differentiate between normal and pathological voices with respect to Parkinson's Disease (PD). The voice signals are extracted using two speech features which are phonation and cepstral features. Phonation features depend on vibrations of vocal folds which involve the stability of pitch frequency and energy. Meanwhile, the cepstral features are extracted by using MFCC. The neural network is used as a classifier and it has 2 layers and 1 hidden layer. This work utilizes 45 PD samples and 45 normal samples based on Parkinson disease movement disorder society's (PDMDS) database. The accuracy achieved using phonation features is 98% and the accuracy of cepstral features is 81.1%. Since the number of normal and abnormal samples is limited, the performance of this system may degrade should bigger database is tested.

A voice pathology detection system with ANN and SVM techniques for classifying voice signal is proposed in [32]. In this work, Particle Swarm Optimization (PSO) is used to find a global optimum selected parameters for ANN and SVM classifiers. From each voice signal, 3 different types of features are generated, namely acoustic features, common signal features, and noise features. SVD is used and it is divided into 3 groups with the same number of samples, D1, D2, and D3. D1 contains the vowel /a/ uttered in a normal pitch, while D2 includes the sentences. On the other hand, D3 refers to the recorded sentences. The highest accuracy that can be achieved by SVM is 92.77% based on D3 group. Meanwhile, the highest accuracy of ANN is 93.27% which is also based on D3 group. Despite the high accuracy, this system has not been evaluated in terms of other vowels such as /u/ and /i/ which are uttered in different intonations in SVD.

A system that is established on the Linear Prediction (LP) analysis to differentiate between healthy and disordered voice

samples is discussed in [33]. In this system, the vocal tract is considered as a linear tube where it is divided into a number of tubes based on the order of LP analysis. GMM algorithm is used to classify voice signals with a different number of Gaussian mixtures such as 4, 8, 16, 32, and 50, where the accuracy improves when the number of Gaussian mixtures increases. Healthy and pathological voice samples are taken from MEEI database with two types of voice data, vowel /a/ and sentences, where there are 173 pathological voice samples with 70 males and 103 females. The number of healthy samples is 53 with 21 males and 32 females. The accuracy results for the vowel /a/ and sentences are 99.94% and 99.75%, respectively. Similar to other above-mentioned works, this system is only evaluated according to vowel /a/, and normal samples are limited too.

A new noise detection method is presented to analyze and differentiate pathological voices from healthy voices in [34]. This method uses a clustering algorithm named Density-Based Spatial Clustering of Applications with Noise (DBSCAN). DBSCAN has many advantages, where it works efficiently even in a big database, and it has the ability to detect noises and clusters of arbitrary shapes. In this method, Mahalanobis distance is adopted to compute the distance between the new object and the different means of clusters and the points which are identified as noises by DBSCAN. The MFCC is then applied to extract features from voice samples. Subsequently, the output features are fed to the SVM classifier to differentiate pathological voices from normal voices. All samples are taken from MEEI database with 53 normal voices and 173 pathological voices that include diseases such as vocal polyp, adductor, keratosis, and paralysis. The classification accuracy based on MEEI database using DBSCAN-SVM can reach up to 98%. However, the database used in this work suffers some limitations, which include different recording environments for pathological and normal voices, and the voice signals are sampled in various frequencies. Additionally, the performance of DBSCAN-SVM in this work is only evaluated by the accuracy.

An automatic voice evaluation system using four different feature selection algorithms is elaborated in [35]. The algorithms that are used to analyze and distinguish voices as healthy or pathological are mRMR (Minimum Redundancy Maximum Relevance), PCA (Principal Component Analysis), Relief algorithm, and LDA (Linear Discriminant Algorithm). These algorithms are applied to eliminate the frequently appearing features in order to reduce the dimension of the original feature. In addition, there are other different techniques of feature extraction used in this system such as Recurrence Period Density Entropy (RPDE), Detrended Fluctuation Analysis, and MFCC. In this system, SVM and ELM techniques are used to classify voice signals. Voice samples are taken from the People's Liberation Army General Hospital located in China, where there are 200 healthy samples and 605 disordered samples. The maximum accuracy of SVM is 77.55%, while the maximum accuracy of ELM is 80.58% by using LDA feature selection. Nonetheless, the system is evaluated based on the vowel /a/ only, and there is no evaluation in terms of sentences or other vowels.

Another voice analysis system which is based on an optimized MFCC is proposed in [36]. The system utilizes 13 original MFCCs as feature extraction and combined MFCC derivatives to discriminate between pathological and normal voices. The LDA feature reduction technique is applied to reduce vector dimensionality of the resulting MFCC. The ANN algorithm is used as a classifier of the proposed work. The voice samples are selected from SVD with 50 healthy and 70 pathological samples of the sustained vowel /a/. Four different types of pathologies are tested, which are Reinke edema with 19 samples, Chronical laryngitis with 24 samples, 21 samples of Spasmodic dysphonia, and 6 samples of Cyst. These pathologies are considered due to their widespread illnesses and tricky medical examinations. The highest accuracy achieved by ANN in this system is 87.82%, and the total number of voice samples is still considered small for accurate performance evaluation.

From the above-mentioned recent studies on various methods for the detection of voice pathology, we can observe some limitations of these proposed works. These limitations can be summarized as follows:

- Most studies ignore pathology classifications and focus on pathology detections only.
- Only the vowel /a/ is utilized in the voice database, while other vowels and sentences are ignored.
- Most proposed works consider a small size of database for healthy and pathological voices.
- Some studies still suffer low accuracy and take a long execution time.

Taking these limitations into account, our proposed work involves the following aspects:

- The detection and classification of voice pathology.
- The samples include sentences and three different vowels with respect to their pitches.
- The database includes an expansive number of healthy and pathological voices.
- Propose OSELM classification algorithm in the voice pathology application, where this algorithm has been considered as an accurate and a fast classifier [37].

In healthcare, the situation is so critical where the system must use an effective classifier such as OSELM algorithm to detect and classify the voice signal accurately and rapidly. This algorithm supports online applications where it can learn data one-by-one or chunk-by-chunk and discards the data in which the training has already been done. On the contrary, batch learning algorithms such as SVM, CNN and ELM use the past data together with the new data whenever new data is received and perform retraining, thus consuming a lot of time.

Therefore, online sequential learning algorithms such as OSELM are preferred over batch learning algorithms as sequential learning algorithms do not require retraining

whenever new data is received. This is considered a significant advantage for OSELM to be used in the application such as voice pathology detection and classification. Table 1 shows advantages and disadvantages of OSELM and other algorithms. In [38] the performance of the OSELM algorithm has been evaluated and compared with other well-known sequential learning algorithms such as Resource Allocation Network (RAN), Growing And Pruning- Radial Basis Function (GAP-RBF), RAN Extended Kalman Filter (RANEKF), Generalized GAP-RBF (GGAP-RBF), Minimal RAN (MRAN), and Stochastic Gradient Descent Back-Propagation (SGBP). The obtained results have shown that OSELM algorithm achieved the highest accuracy with lower time as compared with other sequential learning algorithms.

III. THE PROPOSED METHODOLOGY

In this work, we propose a voice pathology detection and classification system using OSELM technique in which the algorithm has been demonstrated to be extremely fast with good generalization performance [37]. Fig. 1 shows the proposed system with three main phases. The first phase involves voice samples from the database, the second phase includes feature extraction of the voice signals, and the third phase is about detection and classification. These three phases will be explained in the following subsections respectively.

A. THE VOICE DATABASE

SVD (Saarbrucken Voice Database) was used in our experiments as it contains a collection of voice recordings from more than 2000 persons. SVD has recordings of 687 healthy voices from 259 males and 428 females, and 1354 pathological voices from 627 males and 727 females with more than 71 different pathologies. All voice recordings are sampled at frequency of 50 kHz with a resolution of 16-bit. The duration of voice samples is between 1 and 3 sec. The average speakers' age is above 15 years old.

SVD consists of two different recordings for voice samples; the first one refers to recordings of vowels /a/, /u/, and /i/, which are uttered in different intonations. For instance, the vowel ''a'' is uttered as /a/_neutral (a_n), /a/_high (a_h), and /a/_low (a_l), and this is the case for other vowels as well. The second recordings refer to continuous recording sentences in German. For example, ''Guten Morgen, wie geht es Ihnen?'' which means ''Good morning, how are you?''. For our experiment, we categorized all voice samples obtained from SVD into two groups. The first group included all healthy and pathological voice samples in SVD. It involved all sustained vowels in different intonations and all continuous sentences. For the second group, the voice samples were from three types of pathologies namely vocal-fold cyst, vocal-fold polyp, and vocal-fold paralysis. These pathologies are considered common and widely used in the voice pathology systems. The voice samples for these three pathologies have been selected with the vowel /a/ that was uttered in normal or neutral intonation. This was done in order to detect healthy and pathological voices, as well as to classify between three different pathologies. Furthermore, we allocated 80% of voice samples for training and the remaining 20% for testing process.

B. MEL-FREQUENCY CEPSTRAL COEFFICIENT (MFCC)

MFCC is a feature extraction technique that is widely used in automatic speech and speaker recognition. It is based on the auditory system of the human peripheral. The human perception of the contents of sound frequency for speech signals does not follow a linear scale. Thus, for each tone with an actual frequency measured in Hz, a subjective pitch is measured on a scale called the 'Mel Scale' [39]. The Mel frequency scale is a linear frequency spacing below 1000 Hz and logarithmic spacing above 1 kHz. The design of the MFCC technique includes processing the input audio signal. There are different processes performed on the input signals such as Pre-emphasis, Framing, Hamming windowing, Fast Fourier Transform, Mel-Filter bank, and Discrete Cosine Transform. Fig. 2 shows the diagram of feature extraction processes based on MFCC.

In the preprocessing phase, the speech signal processing should be done prior to any other processes. It includes the conversion of analog signal to digital signal, sampling, and quantization. The pre-emphasis step is then performed by passing the voice signal through a filter that emphasizes higher frequencies. In other words, this step increases the signal energy at a higher frequency as the following equation:

$$S'_n = S_n - 0.95 * S_{n-1} \tag{1}$$

where: S'_n is the new sample value, S is the sample value, and n refers to the sample number. The frame step is performed by separating utterance into frames. Each frame has a size, $Tw = 25$ ms, and the number of milliseconds between the left edges of successive windows is called frame shift, $Ts = 10$ ms. The frame size in samples, Nw, can be computed as follows:

$$Nw = Tw \text{ x } Sampling\ rate \tag{2}$$

Therefore, for sampling rate of 44100 samples/second, the frame size is calculated as 1103 samples. Meanwhile, frame shift in samples, Ns, can be calculated as follows:

$$Ns = Ts \text{ x } Sampling\ rate \tag{3}$$

Hence, for this work, Ns is given as 441. After utterance is separated into frames, all frames that contain only zeroth elements are deleted. Then, a Hamming window is applied on each frame of the utterance by considering the next block of feature extraction processing chain and consolidating all the closest frequency lines. The equation of the Hamming window is as follows:

$$W_n = 0.54 - 0.46cos(\frac{2\pi n}{Nw}) \tag{4}$$

where n is sample number and Nw refers to the frame size in samples. Fast Fourier Transform (FFT) is employed to convert each frame of samples from the time domain into the

TABLE 1. Comparison between algorithms.

Techniques	Types	Advantages	Disadvantages
CNN	Offline	1. Automatically extract important features without any human supervision. 2. CNN is also computationally efficient. 3. It provides an efficient dense network that performs the prediction efficiently	1. CNN do not encode the position and orientation of the object into their predictions. 2. Lack of ability to be spatially invariant to the input data. 3. The CNN requires a big database. 4. It retrains the whole database when there is new data which leads to consuming more time.
ELM	Offline	1. Short training time, where it needs less training time compared to other algorithms. 2. Ease of implementation. 3. Minimal human intervention.	1. Over-fitting problem (i.e., many hidden layer nodes). 2. The values of the input layer and the output layer are generated randomly. 3. The ELM retrains the whole database when there is new data which leads to consuming more time.
Decision Tree (DT)	Offline	1. Easy to understand (i.e., explainable algorithm). 2. Fast to prediction. 3. Order of instances has no effect.	1. Classes must be mutually exclusive (i.e., non-overlapping). 2. Missing values of an attribute. 3. Less accurate because the attribute and the class frequencies affect the accuracy.
SVM	Offline	1. SVM works relatively well for the data that is not regularly distributed and have unknown distribution. 2. SVM is more effective in high dimensional spaces. 3. SVM is relatively memory efficient.	1. SVM algorithm is not suitable for large data sets. 2. SVM does not perform well when the data set has more noise (i.e. target classes are overlapping). 3. It needs a long training time.
GMM	Offline	1. GMM can be extended to multiple dimensions. 2. It is flexible in terms of cluster covariance. 3. GMM works well with overlapping clusters.	1. GMM can be sensitive to normalizations of the model or parameters. 2. The performance of GMM is inefficiency in small samples. 3. GMM is not considered scalable.
OSELM	Online	1. It supports online applications (online learning algorithm). 2. When there is new data, the OSELM will train the new data only (no time consumption). 3. It provides better generalization performance at a much faster learning speed.	1. In OSELM, the values of input weights for the hidden layer are generated randomly and not optimum.

FIGURE 1. Flowchart of the proposed system in voice pathology detection and classification.

frequency domain. To perform FFT, the input length is the next power of 2 from Nw, that is $Nw < 2^{p} \longrightarrow 1103 < 2^{11}$. Hence, the length of FFT is $nfft = 2048$.

The voice signal consists of tones with various frequencies. As mentioned previously, every tone with an actual frequency, f, measured in Hz, is a subjective pitch that is measured on the Mel scale. Each filter has three cut-off frequencies with a triangular shape and equal to unity at the center frequency and decrease linearly to zero at the center frequency of two adjacent filters. Then, each filter output is the sum of its filtered spectral components. The Mel-frequency scale is a linear frequency spacing below 1000Hz and a logarithmic spacing above 1000Hz. The following equation is used to convert frequencies from (Hertz) to (Mel):

$$f_{mel} = 2595 \times \log 10\left(1 + \frac{f_{hz}}{700}\right) \tag{5}$$

The last step is the Discrete Cosine Transform (DCT) process to convert the log Mel spectrum into the time domain. The result of the conversion is called Mel-frequency cepstral coefficient. The set of coefficients is called acoustic vectors. Finally, each input utterance is transformed into a sequence of acoustic vectors.

C. ONLINE SEQUENTIAL EXTREME LEARNING MACHINE (OSELM)

The ELM is an offline supervised batch learning algorithm that requires the availability of all data samples in order to perform the training process [40], [41]. However, all data are not available at once in the realistic application, where data are collected in packets over time. Due to this fact, an improved version of ELM, called OSELM has been proposed in [42] to overcome this issue. The OSELM algorithm aims to deal with the emerging needs in several applications

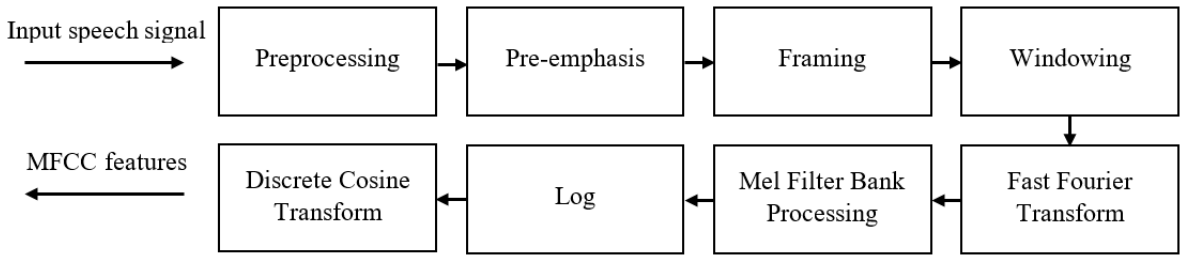

FIGURE 2. The MFCC Diagram.

of online learning. OSELM is considered as a fast algorithm and it is preferred over other algorithms because OSELM eliminates the retraining step upon receiving new data. OSELM is able to learn from the training data through a chunk-by-chunk mechanism with constant or varying length. More details of OSELM algorithm can be found in [42]. In the OSELM algorithm, there are three layers or nodes which are input layer, hidden layer, and output layer. The input layer has the extracted features, the hidden layer has biases, and the output layer has the final classes of the algorithm. The output matrix (H) of the hidden layer is calculated as the following equation:

$$H = W_1 \cdot X_1 + B_1 \tag{6}$$

where W indicates the input weights that link the input layer to the hidden layer, X refers to extracted features by MFCC in the input layer, and B indicates biases of the hidden layer. The input weights (W) and hidden biases (B) are randomly generated with a range between -1 and 1. For N arbitrary distinct samples (x_j, t_j), where $x_j \in \mathrm{R}^{\mathrm{d}}$, and $t_j \in \mathrm{R}^{\mathrm{m}}$, Single Layer Feedforward Neural Networks (SLFNs) with n hidden nodes and the activation function $g(x)$ can be mathematically modeled as the following equation:

$$f\,(\mathrm{X}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \beta_i \mathrm{g}\left(\omega_\mathrm{i} \cdot \mathrm{x_j} + \mathrm{b_i}\right) = \mathrm{t_j}, \quad \mathrm{j} = 1, 2, \ldots, \mathrm{N} \tag{7}$$

Also, equation (7) can be compacted and rewritten as follows:

$$H\beta = T \tag{8}$$

where:

$$H = \begin{pmatrix} g\left(\omega_1 \cdot x_1 + b_1\right) & \ldots & g\left(\omega_n \cdot x_1 + b_n\right) \\ \vdots & \ddots & \vdots \\ g\left(\omega_1 \cdot x_N + b_1\right) & \cdots & g\left(\omega_n \cdot x_N + b_n\right) \end{pmatrix}_{N \times n},$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_n^T \end{bmatrix}_{n \times m}, T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m}$$

The output weights ($\hat{\beta}$) is then estimated according to the following equation:

$$\hat{\beta} = H^{\dagger} T \tag{9}$$

where $H^{\dagger}$ is the Moore-Penrose generalized inverse (pseudo inverse) of the hidden layer output matrix H, and it is calculated as follows:

$$H^{+} = \left(H^T H\right)^{-1} H^T \tag{10}$$

OSELM is executed to learn the training samples successively and incrementally. The learning process of OSELM consists of two steps, initialization step and sequential learning step. In the initialization step, the output matrix of the hidden layer H_0 and the output weights of the initial β_0 are calculated as the equations below:

$$H_{k+1} = g\left(W \cdot X_{k+1} + B\right) \tag{11}$$

$$P_0 = \left(H_0^T H_0\right)^{-1} \tag{12}$$

$$\beta_0 = P_0 H_0^T T_0 \tag{13}$$

In the sequential learning step, the output matrix of the hidden layer H_{k+1} will be updated for the new sample as shown in equation (11). Furthermore, the output weight matrix β_{k+1} will be updated according to the following equations:

$$P_{k+1} = P_k - P_k H_{k+1}^T \left(I + H_{k+1} P_k H_{k+1}^T\right)^{-1} H_{k+1} P_k \tag{14}$$

$$\beta_{k+1} = \beta_k + P_{k+1} H_{k+1}^T \left(T_{k+1} - H_{k+1} \beta^k\right) \tag{15}$$

The set $= k + 1$ and go back to equations (11), (14), and (15) to train the next sample. When all samples are trained, the OSELM can be used for prediction of an unknown input vector. In the OSELM algorithm, the input layer is implemented randomly before further calculation is performed to obtain the output layer and the final results. Fig. 3 shows the architecture of OSELM algorithm, where the final classes are labelled as T_0 and T_1 which refer to pathological and healthy voices, respectively.

The OSELM has two major phases. The boosting phase trains the SLFNs utilizing ELM technique with some training data in the initialization phase, followed by discarding these boosting training data when the boosting stage is finished. After the boosting stage, the OSELM learns the training data chunk by chunk, and discards all the training data when the process of the data learning is finished.

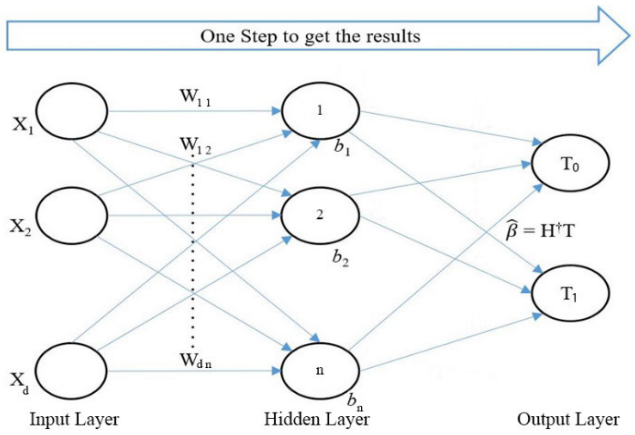

FIGURE 3. OSELM architecture.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In our experiment, the voice samples were taken from the SVD. In order to precisely evaluate the OSELM algorithm in the voice pathology field, we used all sentences and vowels /a/, /i/, and /u/ which were produced in different pitch intonations (i.e., normal, low, and high). The features of these vowels are very important to show the performance of the human larynx, and these voices come out directly from the larynx without the need to use mouth organs. From these two recording sessions, this work made use of 687 healthy voices consisting of 259 males and 428 females, and 1354 pathological voices consisting of 627 males and 727 females. The voices included more than 71 different pathologies. The duration of the voice signals is 1 sec and the frame length is 25 ms. In addition, the total number of extracted features from the voice signal is 2210. In the experiment, the OSELM was implemented with a different number of hidden nodes in the range 50–950 (i.e., the experiment started at 50 nodes and finished at 950 nodes) with the increment step of 50 nodes. Therefore, the total number of experiments was 19. About 80% of the total database was used for data training and the remaining 20% was used for data testing.

All experiments were conducted using Python 3 programming language, where the training and the testing processes were performed using Google Colaboratory server over a PC of 2.50 GHz with 6 GB RAM and HDD 1 TB. The performance of OSELM algorithm was evaluated with widely used measures which are accuracy, precision, recall (sensitivity), F-measures, G-mean, specificity, and execution time as shown in equation (16) to equation (21) [43]–[45]. The evaluation measurements in the equations are described as follows:

- TP (True Positive): the voice is pathological and the algorithm has differentiated it as pathological.
- TN (True Negative): the voice is healthy and the algorithm has differentiated it as healthy.
- FP (False Positive): the voice is healthy whereas the algorithm has differentiated it as pathological.
- FN (False Negative): the voice is pathological whereas the algorithm has differentiated it as healthy.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{16}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{17}$$

$$\text{Recall (Sensitivity)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{18}$$

$$\text{F} - \text{measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Recall} + \text{Precision}} \tag{19}$$

$$\text{G} - \text{Mean} = \sqrt[2]{\frac{\text{tp}}{\text{p}} \times \frac{\text{tn}}{\text{n}}} \tag{20}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{21}$$

Based on the experiments, the results have shown that the OSELM algorithm has the ability to differentiate healthy voices from pathological voices in all vowels with different pitches as well as in sentences from the SVD. From the best results obtained, the accuracy is 91.17% and the G-mean is 87.55% for /a_h/ database at hidden nodes of 900. The precision achieved is 94% for /a_l/ at hidden nodes of 300, and the recall achieved is 91% at hidden nodes of 350 by all /a/ database, where vowels a_h, a_l, and a_n are combined. Moreover, F-measure obtained is 87% for all /i/ when the hidden nodes are 900 and specificity attained is 97.67% for /a_l/ for 300 hidden nodes. The best time for training and testing OSELM classifier is 0.70 sec for /u_n/ database when the hidden nodes are 750. Fig. 4 shows the best results achieved by OSELM algorithm in the voice pathology detection. However, the minimum accuracy obtained is 77.21% for all sentences and vowels in SVD. This is due to the fact that OSELM algorithm faces difficulties to classify voices with different vowels and sentences which result in less accuracy. The overall results of OSELM algorithm for the first group from SVD are shown in Table 2. The first group includes all healthy and pathological voices that involve all sustained vowels in different intonations and all continuous sentences. From the obtained results, it can be observed that OSELM algorithm performs well in terms of training and classifying voice signals when the hidden nodes are at 300 for precision and specificity, 350 for recall, 750 for execution time, and 900 for accuracy, F-measure, and G-mean.

The accuracy is widely used as a significant measure to evaluate the voice pathology systems. Therefore, we have chosen this performance measurement in this study in order to compare the proposed OSELM model with other methods which have also utilized voice samples from SVD in the voice pathology systems. The comparison is made in terms of running sentences and all sustained vowels, /a/, /i/, and /u/, that are produced in high, low, and normal pitches. We compare our proposed method with the work in [46] for all vowels with respect to the three pitches, and with the method in [47] for /a_h/ pitch. In addition, we compare the OSELM algorithm with the method in [48] for /a_h/

TABLE 2. The results of OSELM in SVD.

Hidden nodes	Accuracy	Precision	Recall	F-measure	G-mean	Specificity	Time (sec)
a_h							
900	**91.17%**	89%	80%	84%	**87.55%**	95.83%	1.10
a_l							
300	87%	**94%**	67%	78%	80.69%	**97.67%**	0.99
a_n							
850	88.15%	78%	74%	76%	82.77%	93%	0.84
All_a							
350	81%	76%	**91%**	82%	80%	70%	1.19
Hidden nodes	Accuracy	Precision	Recall	F-measure	G-mean	Specificity	Time (sec)
i_h							
350	82%	83%	60%	70%	75%	94%	0.96
i_l							
400	82.08%	81%	68%	74%	78.43%	90.47%	1.07
i_n							
650	84%	81%	75%	78%	82%	89.13%	1.17
All_i							
900	82%	90%	85%	**87%**	78%	72%	1.27
Hidden nodes	Accuracy	Precision	Recall	F-measure	G-mean	Specificity	Time (sec)
u_h							
500	81%	71%	68%	70%	77%	87%	0.84
u_l							
850	84%	70%	86.36%	78%	84.62%	82.22%	0.83
u_n							
750	87%	83%	76%	79%	84%	92%	**0.70**
All_u							
650	81%	88%	75%	81%	81.26%	88%	1.31
Hidden nodes	Accuracy	Precision	Recall	F-measure	G-mean	Specificity	Time (sec)
Sentences							
500	81.33%	75%	84.37%	79%	81.67%	79.06%	0.98
Hidden nodes	Accuracy	Precision	Recall	F-measure	G-mean	Specificity	Time (sec)
All Database							
700	77.21%	88%	73%	80%	78.20%	84%	1.03

and /u_h/ pitches. Furthermore, we also compare between OSELM algorithm and the studies in [49]–[52], where the works in [49]–[51] were based on /a_n/ pitch only, and the work in [52] utilized sentences and all vowels pronounced in normal pitch. Table 3 shows the comparison of accuracy between OSELM and other methods in all sessions of the SVD. Results show that our work using OSELM in the voice pathology detection system outperforms all other methods in terms of accuracy.

Most systems of voice pathology detection and classification were evaluated in terms of accuracy, specificity, and sensitivity only. There has been not much evaluation conducted based on execution time in the literature review. Hence, the proposed OSELM is also compared with the method in [53] in terms of the execution time. The method in [53] utilized the CNN algorithm in voice pathology detection using the vowel /a/ that was produced in normal pitch. The best execution time for CNN is 2.54 sec for testing the voice signal. On the other hand, the execution time for the proposed OSELM is 0.84 sec where it outperforms CNN by 66.93%. Table 4 shows the comparison of execution time between OSELM and CNN based on the vowel /a_n/.

For the second set of experiments, OSELM algorithm is evaluated in terms of detection and classification of three common types of pathologies namely vocal-fold cyst, vocal-fold polyp, and vocal-fold paralysis. The detection and classification are based on voice samples with vowel /a/ uttered in normal or neutral intonation. From the results, the highest accuracy achieved is 89.47% with the hidden nodes of 150 for the detection between healthy and three pathological voices. Meanwhile, the highest achieved accuracy for the classification between cyst versus polyp and paralysis, paralysis versus polyp and cyst, and polyp versus paralysis and cyst are 97.72%, 88.9%, and 88.8%, respectively. Table 5 presents the overall results for voice pathology detection and classification based on these three types of pathologies. Table 6 lists the comparison of the highest accuracy achieved between OSELM algorithm and other methods using the SVD for the detection and classification of voice pathology with respect to cyst, polyp, and paralysis pathologies.

Furthermore, we compare the performance of OSELM with SVM in [57] in terms of the execution time in the classification of voice pathologies. The best execution time

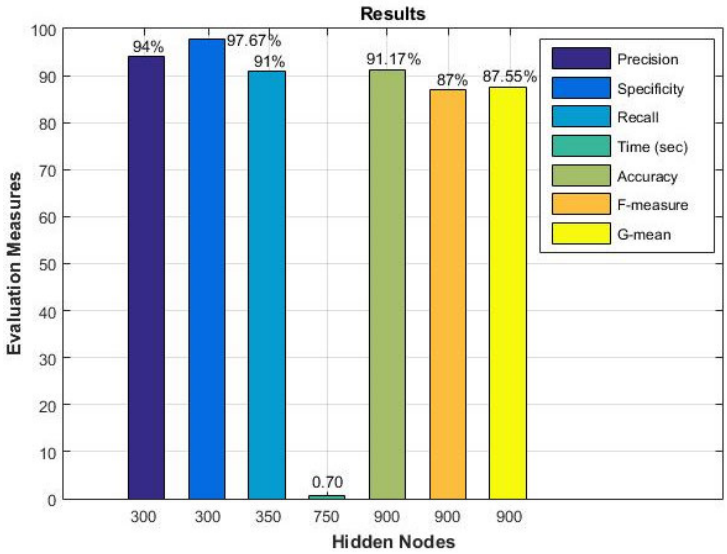

FIGURE 4. Best achieved results by OSELM classifier.

TABLE 3. Comparison of accuracy between OSELM and other methods.

Methods	Accuracy	Methods	Accuracy	Methods	Accuracy	Methods	Accuracy
a_h		**a_n**		**a_l**		**All_a**	
OSELM	**91.17%**	OSELM	**88.15%**	OSELM	**87%**	OSELM	**81%**
		GMM [46]	67%				
GMM [46]	66.6%	DNN [49]	68.08%				
		SVM [50]	85.77%				
SVM [47]	71.12%	DNN [51]	82.01%	GMM [46]	65.6%	GMM [46]	71.8%
CNN [48]	73%	SVM [52]	74.32%				
Methods	Accuracy	Methods	Accuracy	Methods	Accuracy	Methods	Accuracy
i_h		**i_n**		**i_l**		**All_i**	
OSELM	**82%**	OSELM	**84%**	OSELM	**82.08%**	OSELM	**82%**
		GMM [46]	64.5%				
GMM [46]	64%	SVM [52]	72.29%	GMM [46]	64.2%	GMM [46]	71%
Methods	Accuracy	Methods	Accuracy	Methods	Accuracy	Methods	Accuracy
u_h		**u_n**		**u_l**		**All_u**	
OSELM	**81%**	OSELM	**87%**	OSELM	**84%**	OSELM	**81%**
GMM [46]	64%	GMM [46]	63.4%				
CNN [48]	63%	SVM [52]	71.45%	GMM [46]	64.6%	GMM [46]	71.5%
Methods				Accuracy			
Sentences							
OSELM				**81.33%**			
SVM [52]				76.19			

TABLE 4. Comparison of execution time between OSELM and CNN.

Methods	Execution Time (sec)
OSELM	**0.84**
CNN [53]	2.54

for SVM is 97.48 sec compared to OSELM which completed computation within 0.63 sec only. Therefore, this clearly shows that OSELM is able to deliver rapid classification results compared to other well-known algorithms. Table 7 shows the comparison of execution time between OSELM and SVM.

Despite the encouraging results obtained in the detection and classification of voice pathology, OSELM algorithm has some limitations which can be summarized as follows: (i) the input weights of OSELM are generated randomly in which it results in inconsistency of the accuracy, and (ii) this algorithm has been trained and tested in the voice pathology system based on SVD only.

TABLE 5. The detection and classification results for three pathologies.

Detection							
Hidden nodes	Accuracy	Precision	Recall	F-measure	G-mean	Specificity	Time (sec)
150	89.47%	82%	100%	90%	89.44%	80%	0.77
Classification							
Hidden nodes	Accuracy	Precision	Recall	F-measure	G-mean	Specificity	Time (sec)
Cyst vs. (Polyp and Paralysis)							
250	97.72%	100%	100%	99%	98.85%	97.7%	0.80
Paralysis vs. (Polyp and Cyst)							
300	88.9%	87.5%	100%	93%	70.71%	50%	0.78
Polyp vs. (Paralysis and Cyst)							
350	88.8%	50%	100%	67%	93.54%	87.50%	0.63

TABLE 6. Comparison of accuracy between methods based on three pathologies.

Methods	Accuracy
Detection	
OSELM	**89.47%**
DNN [54]	87.4%
GMM [55]	80.2%
Cyst vs. (Polyp and Paralysis)	
OSELM	**97.72%**
SVM [56]	97.5%
Paralysis vs. (Polyp and Cyst)	
OSELM	**88.9%**
SVM [56]	79.17%
Polyp vs. (Paralysis and Cyst)	
OSELM	**88.8%**
SVM [56]	82.08%

TABLE 7. Comparison of execution time between OSELM and SVM.

Methods	Execution Time (sec)
OSELM	**0.63**
SVM [57]	97.48

V. CONCLUSION AND FUTURE WORK

This study has presented the OSELM algorithm for the detection and classification of voice pathology based on voice samples from SVD. In this study, we have utilized two types of voice recordings which are vowels /a/, /i/, and /u/ produced at high, low, and normal pitches, as well as continuous sentences. The algorithm is also tested for detection and classification of three types of pathologies which are cyst, polyp, and paralysis. The experimental results demonstrate that the OSELM algorithm is able to differentiate healthy and pathological voices with maximum accuracy of 91.17%, 94% of precision, and 91% of recall. Moreover, the maximum results achieved for F-measure, G-mean, and specificity are 87%, 87.55%, and 97.67%, respectively. The best achieved accuracy for detection and classification of three pathologies are 89.47% and 97.72%, respectively. It is worth to mention that this is the first work that is based on OSELM algorithm in the voice pathology detection and classification, and this algorithm is proven to be fast and accurate in the acoustic systems. Future work includes improvements in the OSELM classifier such as tuning OSELM by choosing the optimal weights. Furthermore, based on its potential, OSELM algorithm can be further tested for detection and diagnosis of particular diseases of the voice box with respect to real-time clinical applications.

ACKNOWLEDGMENT

The authors would like to thank the Ministry of Higher Education (MOHE), Research Management Centre of Universiti Teknologi Malaysia (UTM) and School of Electrical Engineering, UTM, and Universiti Kebangsaan Malaysia (UKM) for the sponsorship and collaboration for this research.

REFERENCES

[1] J. Morawska and E. N. Bogusz, "Risk factors and prevalence of voice disorders in different occupational groups—A review of literature," *Otorynolaryngologia PrzegląD Kliniczny*, vol. 16, no. 3, pp. 94–102, 2017.

[2] *National Institute on Deafness and Other Communication Disorders. Voice, Speech, and Language, USA*. Accessed: Jul. 15, 2020. [Online]. Available: https://www.nidcd.nih.gov/health/statistics

[3] A. Al-nasheri, G. Muhammad, M. Alsulaiman, and Z. Ali, "Investigation of voice pathology detection and classification on different frequency regions using correlation functions," *J. Voice*, vol. 31, no. 1, pp. 3–15, Jan. 2017.

[4] A. Al-Nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, K. H. Malki, T. A. Mesallam, and M. F. Ibrahim, "Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions," *IEEE Access*, vol. 6, pp. 6961–6974, 2018, doi: 10.1109/ACCESS.2017.2696056.

[5] A. Ali and S. Ganar, "Intelligent pathological voice detection," *Int. J. Innov. Res. Technol.*, vol. 5, no. 5, pp. 92–95, Oct. 2018.

[6] O. Saz, J. Simón, W. R. Rodríguez, E. Lleida, and C. Vaquero, "Analysis of acoustic features in speakers with cognitive disorders and speech impairments," *EURASIP J. Adv. Signal Process.*, vol. 2009, no. 1, pp. 1–11, Dec. 2009.

[7] F. T. Al-Dhief, N. M. A. Latiff, N. N. N. A. Malik, N. S. Salim, M. M. Baki, M. A. A. Albadr, and M. A. Mohammed, "A survey of voice pathology surveillance systems based on Internet of Things and machine learning algorithms," *IEEE Access*, vol. 8, pp. 64514–64533, 2020, doi: 10.1109/ACCESS.2020.2984925.

[8] P. Song, "Assessment of vocal cord function and voice disorders," in *Principles and Practice of Interventional Pulmonology*. New York, NY, USA: Springer, 2013, pp. 137–149.

[9] T. Drugman, T. Dubuisson, and T. Dutoit, "On the mutual information between source and filter contributions for voice pathology detection," 2020, *arXiv:2001.00583*. [Online]. Available: http://arxiv.org/abs/2001.00583
[10] C. Gadepalli, F. Jalalinajafabadi, Z. Xie, B. M. G. Cheetham, and J. J. Homer, "Voice quality assessment by simulating GRBAS scoring," in *Proc. Eur. Model. Symp. (EMS)*, Nov. 2017, pp. 107–111, doi: 10.1109/EMS.2017.29.
[11] W. Yuanbo, Z. Changwei, F. Ziqi, Z. Yihua, Z. Xiaojun, and T. Zhi, "Voice pathology detection and multi-classification using machine learning classifiers," in *Proc. Int. Conf. Sens., Meas. Data Anal. Era Artif. Intell. (ICSMD)*, Oct. 2020, pp. 319–324, doi: 10.1109/ICSMD50554.2020.9261710.
[12] O. I. Obaid, M. A. Mohammed, M. K. A. Ghani, A. Mostafa, and F. Taha, "Evaluating the performance of machine learning techniques in the classification of Wisconsin breast cancer," *Int. J. Eng. Technol.*, vol. 7, no. 4, pp. 160–166, 2018.
[13] E. H. Ahmed, M. R. M. ALsemawi, M. H. Mutar, H. O. Hanoosh, A. H. Abbas, F. T. AL-Dhief, and M. A. A. Albadr, "Convolutional neural network for the detection of coronavirus (COVID-19) based on X-ray images," *Solid State Technol.*, vol. 63, no. 4, pp. 8730–8739, 2020.
[14] M. K. Shahsavari, H. Rashidi, and H. R. Bakhsh, "Efficient classification of Parkinson's disease using extreme learning machine and hybrid particle swarm optimization," in *Proc. 4th Int. Conf. Control, Instrum., Autom. (ICCIA)*, Jan. 2016, pp. 148–154, doi: 10.1109/ICCIAutom.2016.7483152.
[15] S. R. K. Sharma and A. K. Gupta, "Processing and analysis of human voice for assessment of parkinson disease," *J. Med. Imag. Health Informat.*, vol. 6, no. 1, pp. 63–70, Feb. 2016.
[16] I. M. M. El Emary, M. Fezari, and F. Amara, "Towards developing a voice pathologies detection system," *J. Commun. Technol. Electron.*, vol. 59, no. 11, pp. 1280–1288, Nov. 2014.
[17] Y. Maryn, F. Ysenbaert, A. Zarowski, and R. Vanspauwen, "Mobile communication devices, ambient noise, and acoustic voice measures," *J. Voice*, vol. 31, no. 2, pp. 248.e11-248.e23, Mar. 2017.
[18] N. Adiga, V. C. M., K. Pullela, and S. R. M. Prasanna, "Zero frequency filter based analysis of voice disorders," in *Proc. Interspeech*, Aug. 2017, pp. 1824–1828.
[19] P. M. Chauhan and N. P. Desai, "Mel frequency cepstral coefficients (MFCC) based speaker identification in noisy environment using Wiener filter," in *Proc. Int. Conf. Green Comput. Commun. Electr. Eng. (ICGCCEE)*, Mar. 2014, pp. 1–5, doi: 10.1109/ICGCCEE.2014.6921394.
[20] H. Bonifaco, K. R. Guzman, J. N. Jara, A. D. Jasareno, A. C. Zabala, S. V. Prado, and C. S. Buenaventura, "Comparative analysis of filipino-based rhinolalia aperta speech using mel frequency cepstral analysis and Perceptual Linear Prediction," in *Proc. IEEE 9th Int. Conf. Humanoid, Nanotechnol., Inf. Technol., Commun. Control, Environ. Manage. (HNICEM)*, Dec. 2017, pp. 1–6, doi: 10.1109/HNICEM.2017.8269507.
[21] M. M. Baki, G. Wood, M. Alston, P. Ratcliffe, G. Sandhu, J. S. Rubin, and M. A. Birchall, "Reliability of OperaVOX against multidimensional voice program (MDVP)," *Clin. Otolaryngology*, vol. 40, no. 1, pp. 22–28, Jan. 2015.
[22] R. Benhammoud and A. Kacha, "Automatic classification of disordered voices with hidden Markov models," in *Proc. Int. Conf. Signal, Image, Vis. Their Appl. (SIVA)*, Nov. 2018, pp. 1–6, doi: 10.1109/SIVA.2018.8661038.
[23] A. Bajpai, U. Varshney, and D. Dubey, "Performance enhancement of automatic speech recognition system using Euclidean distance comparison and artificial neural network," in *Proc. 3rd Int. Conf. Internet Things, Smart Innov. Usages (IoT-SIU)*, Feb. 2018, pp. 1–5, doi: 10.1109/IoT-SIU.2018.8519839.
[24] T. A. Mesallam, M. Farahat, K. H. Malki, M. Alsulaiman, Z. Ali, A. Al-nasheri, and G. Muhammad, "Development of the arabic voice pathology database and its evaluation by using speech features and machine learning algorithms," *J. Healthcare Eng.*, vol. 2017, pp. 1–13, Oct. 2017.
[25] B. Woldert-Jokisz, "Saarbruecken voice database," Inst. Phonetics, Univ. Saarland, Saarbrücken, Germany, Tech. Rep., 2007.
[26] S.-H. Fang, Y. Tsao, M.-J. Hsiao, J.-Y. Chen, Y.-H. Lai, F.-C. Lin, and C.-T. Wang, "Detection of pathological voice using cepstrum vectors: A deep learning approach," *J. Voice*, vol. 33, no. 5, pp. 634–641, Sep. 2019.
[27] R. R. A. Barreira and L. L. Ling, "Kullback–Leibler divergence and sample skewness for pathological voice quality assessment," *Biomed. Signal Process. Control*, vol. 57, Mar. 2020, Art. no. 101697.
[28] E. S. Fonseca, R. C. Guido, S. B. Junior, H. Dezani, R. R. Gati, and D. C. M. Pereira, "Acoustic investigation of speech pathologies based on the discriminative paraconsistent machine (DPM)," *Biomed. Signal Process. Control*, vol. 55, Jan. 2020, Art. no. 101615.
[29] V. Mittal and R. K. Sharma, "Glottal signal analysis for voice pathology," in *Proc. 2nd Int. Conf. Innov. Electron., Signal Process. Commun. (IESC)*, Mar. 2019, pp. 54–59, doi: 10.1109/IESPC.2019.8902368.
[30] M. Alhussein and G. Muhammad, "Automatic voice pathology monitoring using parallel deep models for smart healthcare," *IEEE Access*, vol. 7, pp. 46474–46479, 2019, doi: 10.1109/ACCESS.2019.2905597.
[31] S. S. Upadhya and A. N. Cheeran, "Discriminating parkinson and healthy people using phonation and cepstral features of speech," *Procedia Comput. Sci.*, vol. 143, pp. 197–202, Jan. 2018.
[32] H. J. A. Laverde, A. E. C. Ospina, and D. H. P. Ordóñez, "Voice pathology detection using artificial neural networks and support vector machines powered by a multicriteria optimization algorithm," in *Proc. Workshop Eng. Appl.*, Cham, Switzerland: Springer, vol. 915, 2018, pp. 148–159.
[33] Z. Ali, G. Muhammad, and M. F. Alhamid, "An automatic health monitoring system for patients suffering from voice complications in smart cities," *IEEE Access*, vol. 5, pp. 3900–3908, 2017, doi: 10.1109/ACCESS.2017.2680467.
[34] R. Amami and A. Smiti, "An incremental method combining density clustering and support vector machines for voice pathology detection," *Comput. Electr. Eng.*, vol. 57, pp. 257–265, Jan. 2017.
[35] Z. Wang, P. Yu, N. Yan, L. Wang, and M. L. Ng, "Automatic assessment of pathological voice quality using multidimensional acoustic analysis based on the GRBAS scale," *J. Signal Process. Syst.*, vol. 82, no. 2, pp. 241–251, Feb. 2016.
[36] N. Souissi and A. Cherif, "Speech recognition system based on short-term cepstral parameters, feature reduction method and artificial neural networks," in *Proc. 2nd Int. Conf. Adv. Technol. Signal Image Process. (ATSIP)*, Mar. 2016, pp. 667–671, doi: 10.1109/ATSIP.2016.7523163.
[37] W. Guo, T. Xu, K. Tang, J. Yu, and S. Chen, "Online sequential extreme learning machine with generalized regularization and adaptive forgetting factor for time-varying system prediction," *Math. Problems Eng.*, vol. 2018, pp. 1–22, May 2018.
[38] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "A fast and accurate online sequential learning algorithm for feedforward networks," *IEEE Trans. Neural Netw.*, vol. 17, no. 6, pp. 1411–1423, Nov. 2006, doi: 10.1109/TNN.2006.880583.
[39] 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, Mar. 2010.
[40] M. A. A. Albadr and S. Tiun, "Spoken language identification based on particle swarm optimisation–extreme learning machine approach," *Circuits, Syst., Signal Process.*, vol. 39, no. 9, pp. 4596–4622, Mar. 2020.
[41] M. A. A. Albadra and S. Tiuna, "Extreme learning machine: A review," *Int. J. Appl. Eng. Res.*, vol. 12, no. 14, pp. 4610–4623, 2017.
[42] G. B. Huang, N. Y. Liang, H. J. Rong, P. Saratchandran, and N. Sundararajan, "On-line sequential extreme learning machine," in *Proc. IASTED Int. Conf. Comput. Intell.*, Jul. 2005, pp. 232–237.
[43] M. A. A. Albadr, S. Tiun, F. T. AL-Dhief, and M. A. M. Sammour, "Spoken language identification based on the enhanced self-adjusting extreme learning machine approach," *PLoS ONE*, vol. 13, pp. 1–27, Apr. 2018.
[44] M. A. A. Albadr, S. Tiun, M. Ayob, and F. T. AL-Dhief, "Spoken language identification based on optimised genetic algorithm–extreme learning machine approach," *Int. J. Speech Technol.*, vol. 22, no. 3, pp. 711–727, Jul. 2019.
[45] M. A. A. Albadr, S. Tiun, M. Ayob, F. T. AL-Dhief, K. Omar, and F. A. Hamzah, "Optimised genetic algorithm-extreme learning machine approach for automatic COVID-19 detection," *PLoS ONE*, vol. 15, no. 12, pp. 1–28, Dec. 2020.
[46] D. Mart?nez, E. Lleida, A. Ortega, A. Miguel, and J. Villalba, "Voice pathology detection on the SaarbrüCken voice database with calibration and fusion of scores using MultiFocal toolkit," in *Advances in Speech and Language Technologies for Iberian Languages*. Berlin, Germany: Springer, vol. 328, 2012, pp. 99–109, doi: 10.1007/978-3-642-35292-811.
[47] A. Al-nasheri, Z. Ali, G. Muhammad, and M. Alsulaiman, "Voice pathology detection using auto-correlation of different filters bank," in *Proc. IEEE/ACS 11th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Nov. 2014, pp. 50–55, doi: 10.1109/AICCSA.2014.7073178.

[48] A. Rueda and S. Krishnan, "Augmenting dysphonia voice using Fourier-based synchrosqueezing transform for a CNN classifier," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 6415–6419, doi: 10.1109/ICASSP.2019.8682391.
[49] P. Harar, J. B. Alonso-Hernandezy, J. Mekyska, Z. Galaz, R. Burget, and Z. Smekal, "Voice pathology detection using deep learning: A preliminary study," in *Proc. Int. Conf. Workshop Bioinspired Intell. (IWOBI)*, Jul. 2017, pp. 1–4.
[50] L. Verde, G. De Pietro, and G. Sannino, "Voice disorder identification by using machine learning techniques," *IEEE Access*, vol. 6, pp. 16246–16255, 2018, doi: 10.1109/ACCESS.2018.2816338.
[51] L. Vavrek, M. Hires, D. Kumar, and P. Drotar, "Deep convolutional neural network for detection of pathological speech," in *Proc. IEEE 19th World Symp. Appl. Mach. Intell. Informat. (SAMI)*, Jan. 2021, Art. no. 000245, doi: 10.1109/SAMI50585.2021.9378656.
[52] S. R. Kadiri and P. Alku, "Analysis and detection of pathological voice using glottal source features," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 2, pp. 367–379, Feb. 2020, doi: 10.1109/JSTSP.2019.2957988.
[53] M. A. Mohammed, K. H. Abdulkareem, S. A. Mostafa, M. K. A. Ghani, M. S. Maashi, B. G. Zapirain, I. Oleagordia, H. Alhakami, and F. T. AL-Dhief, "Voice pathology detection and classification using convolutional neural network model," *Appl. Sci.*, vol. 10, pp. 1–13, May 2020.
[54] J. Moon and S. Kim, "An approach on a combination of higher-order statistics and higher-order differential energy operator for detecting pathological voice with machine learning," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2018, pp. 46–51, doi: 10.1109/ICTC.2018.8539495.
[55] Z. Ali, M. Alsulaiman, G. Muhammad, I. Elamvazuthi, A. Al-Nasheri, T. A. Mesallam, M. Farahat, and K. H. Malki, "Intra- and inter-database study for arabic, english, and German databases: Do conventional speech features detect voice pathology," *J. Voice*, vol. 31, no. 3, pp. 386.e1–386.e8, May 2017.
[56] A. Al-nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, T. A. Mesallam, M. Farahat, K. H. Malki, and M. A. Bencherif, "An investigation of multidimensional voice program parameters in three different databases for voice pathology detection and classification," *J. Voice*, vol. 31, no. 1, pp. 113.e9–113.e18, Jan. 2017.
[57] F. N. C. Kassim, V. Vijean, H. Muthusamy, R. Abdullah, and Z. Abdullah, "Voice pathology analysis using DT-CWPT and ReliefF algorithm," in *Proc. J. Phys., Conf. Ser., Int. Conf. Biomed. Eng.*, Penang Island, Malaysia, vol. 1372, 2019, pp. 1–6.

FAHAD TAHA AL-DHIEF (Graduate Student Member, IEEE) was born in Amarah, Iraq. He received the B.S. degree in software engineering from Imam Ja'afar Al-Sadiq University, Iraq, in 2013, and the M.S. degree in computer science from the Universiti Kebangsaan Malaysia, Malaysia, in 2016. He is currently pursuing the Ph.D. degree with the Department of Communication Engineering, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Malaysia. His research interests include sensor networks, routing protocols, mobile *ad-hoc* networks, social networks, the Internet of Things, machine learning, artificial neural networks, deep learning, and location-based service. He is a member of IEEE Communications Society.

MARINA MAT BAKI graduated from the Faculty of Medicine, Universiti Kebangsaan Malaysia (National University of Malaysia). She received the Ph.D. degree in laryngology from the University College London, in November 2014. In 1997, she started her medical career with the Ministry of Health of Malaysia. In 2002, she started her otorhinolaryngology training with the Universiti Kebangsaan Malaysia Medical Centre (UKMMC). She completed the ORL training four years later in 2006. She has been inspired and trained by Prof. Dato' Dr. Abdullah Sani Mohamed to take up laryngology. During the Ph.D. program, she obtained a certificate of subspeciality training in laryngology, when she worked with distinguished laryngologists and voice surgeons: Prof. Martin Birchall, Dr. John Rubin, and Dr. Guri Sandhu, at the Royal National Throat Nose and Ear Hospital and Charing Cross Hospital, London, U.K., from July 2010 to May 2014. She is currently a Professor of laryngology with the Department of UKMMC. Her research interests include laryngotracheal stenosis, laryngeal paralysis, voice disorders, phonosurgery, and sleep apnoea.

NURUL MU'AZZAH ABDUL LATIFF (Senior Member, IEEE) received the Bachelor of Engineering degree in electrical-telecommunication from Universiti Teknologi Malaysia (UTM), in 2002, and the M.Sc. degree in communication and signal processing and the Ph.D. degree in wireless communication engineering from Newcastle University, in 2003 and 2008, respectively. She is currently a Senior Lecturer with the Division of Communication Engineering, School of Electrical Engineering, (UTM). Her research interests include wireless sensor networks, mobile *ad hoc* networks, cognitive radio, the Internet of Things, and artificial intelligence. She is a member of IEEE Communications Society and Vehicular Technology Society, besides being a certified Chartered Engineer from the IET.

NIK NOORDINI NIK ABD. MALIK (Member, IEEE) received the bachelor's degree in electrical engineering (telecommunication) from the Universiti Teknologi Malaysia (UTM), in 2003, the Master of Engineering (M.Eng.) degree in radio frequency (RF) and microwave communication engineering from The University of Queensland, Australia (UQ), in 2005, and the Doctor of Philosophy (Ph.D.) degree in electrical engineering from the Universiti Teknologi Malaysia (UTM), Malaysia, in 2013. She is currently a Senior Lecturer with the School of Electrical Engineering, (UTM). She worked as a Radio Frequency Research and Development Electrical Engineer with Motorola Technology PDT, Penang, Malaysia. Her research interests include wireless sensor networks, distributed beamforming, and meta-heuristic algorithms.

NASEER SABRI SALIM received the B.Sc. and M.Sc. degrees in computer engineering, with a focus on intelligent computer vision field, and the Ph.D. degree in the development of a novel intelligent WSN. He is currently a Faculty Member with the School of Computer and Communication Engineering, Universiti Malaysia Perlis. He has also authored and coauthored more than 45 technical articles of internationally cited journals. His main research area targeting on building a new intelligent wireless sensor network based on an artificial intelligence paradigm. His research interests include automation, smart wireless sensor network solutions, embedded system design, and computer vision. He is currently serving as a reviewer for various international publications including top-ten ranked journals.

MUSATAFA ABBAS ABBOOD ALBADER received the bachelor's degree from the Department of Computer Science, University of Basrah College of Education, Iraq, in 2011, and the M.S. degree in computer science from the University Kebangsaan Malaysia, Malaysia, in 2017, where he is currently pursuing the Ph.D. degree with the Faculty of Information Science and Technology. His research interests include machine learning, neural networks, speech processing, data mining, parallel processing, and meta-heuristic algorithms.

NOR MUZLIFAH MAHYUDDIN (Member, IEEE) received the B.Eng. degree from Universiti Teknologi Malaysia, Malaysia, in 2005, the M.Sc. degree from the Universiti Sains Malaysia, Malaysia, in 2006, and the Ph.D. degree from Newcastle University, Newcastle upon Tyne, U.K., in 2011. She is currently an Associate Professor and the Doctoral Supervisor with the Universiti Sains Malaysia. Her research interests include RF and microwave engineering, reliability, signal integrity, sparse channel estimation, and OFDM systems. She is a member of IET and a Professional Member of the Association for Computing Machinery (ACM). She is also involved in the Communications Society (ComSoc). She is also registered with the Board of Engineers Malaysia (BEM).

MAZIN ABED MOHAMMED received the B.Sc. degree in computer science from the College of Computer, University of Anbar, Ramadi, Iraq, in 2008, the master's degree in information technology from the Universiti Tenaga Nasional, Malaysia, in 2011, and the Ph.D. degree from the Universiti Teknikal Malaysia Melaka, Malaysia, in 2018. He is currently working as a Teacher with the Planning and Follow Up Department, University of Anbar. He teaches a variety of university courses in computer science, such as operating systems, database design, mobile systems programing, software project management, Web technologies, and software requirements and design. His research interests include artificial intelligence, medical image processing, machine learning, computer vision, computational intelligence, the IoT, biomedical computing, bio-informatics, and fog computing.