

# A Zero-Watermarking Algorithm for Privacy Protection in Biomedical Signals

Zulfiqar Ali <sup>1\*</sup>, Muhammad Imran<sup>2†</sup>, Mansour Alsulaiman<sup>1</sup>, Tanveer Zia<sup>3</sup>,  
Muhammad Shoaib<sup>2</sup>

<sup>1</sup>Digital Speech Processing Group, Department of Computer Engineering, College of Computer and Information Science, King Saud University, Riyadh 11543, Saudi Arabia.

<sup>2</sup>College of Computer and Information Science, King Saud University, Saudi Arabia.

<sup>3</sup>School of Computing and Mathematics, Charles Sturt University, Australia.

[zuali@ksu.edu.sa](mailto:zuali@ksu.edu.sa), [cimran@ksu.edu.sa](mailto:cimran@ksu.edu.sa), [tzia@csu.edu.au](mailto:tzia@csu.edu.au),  
[msuliman@ksu.edu.sa](mailto:msuliman@ksu.edu.sa), [muhshoaib@ksu.edu.sa](mailto:muhshoaib@ksu.edu.sa)

---

## Abstract

Confidentiality of health information is indispensable to protect privacy of an individual. However, recent advances in electronic healthcare systems allow transmission of sensitive information through the Internet, which is prone to various vulnerabilities, attacks and may lead to unauthorized disclosure. Such situations may not only create adverse effects for individuals but may also cause severe consequences such as hefty regulatory fines, bad publicity, legal fees, and forensics. To avoid such predicaments, a privacy protected healthcare system is proposed in this study that protects the identity of an individual as well as detects vocal fold disorders. The privacy of the developed healthcare system is based on the proposed zero-watermarking algorithm, which embeds a watermark in a secret key instead of the signals to avoid the distortion in an audio sample. The identity is protected by the generation of its secret shares through visual cryptography. The generated shares are embedded by finding the patterns into the audio with the application of one-dimensional local binary pattern. The proposed zero-watermarking algorithm is evaluated by using audio samples taken from the Massachusetts Eye and Ear Infirmary voice disorder database. Experimental results demonstrate that the proposed algorithm achieves imperceptibility and is reliable in its extraction of identity. In addition, the proposed algorithm does not affect the results of disorder detection and it is robust against noise attacks of various signal-to-noise ratios.

*Keywords:* E-Healthcare, privacy protection; zero-watermarking; visual cryptography; local binary pattern; MFCC; SVM

---

---

<sup>†</sup> Corresponding author

# 1 Introduction

The idea of privacy protection has long been practiced in healthcare in paper form, in order to hide a patient's information. A simple example is to store a patient's record in lockers. Recent developments in the electronic healthcare (E-health) have contributed significantly by providing effective solutions in the form of healthcare information and diagnostic systems to yield complementary information for an accurate diagnosis as well as treatment plans [1-3]. Health information is classified as sensitive and therefore, protection of this information is always a top priority for patients and healthcare providers [4-6]. The healthcare systems store, manage, and transmit information related to the health of an individual electronically. This information may include personally identifiable information (PII) such as names, dates (birth, admission, discharge, death, and treatment), contact details, social security numbers, medical record numbers, photographs, fingerprints. Unauthorized disclosure of health PII may not only create adverse effects for patients but may also result in hefty regulatory fines, bad publicity, legal fees, and forensics for healthcare providers. For instance, vocal fold disorders degrade the quality of voice, and people involved in voice-related professions such as teachers and lawyers have a high risk of developing such disorders [7]. Unauthorized access to health information of teachers or lawyers may have a negative impact on their careers, especially when performed with malicious intent. Moreover, most countries have very strict policies and regulatory laws that holds healthcare providers accountable for any unauthorized disclosure of such sensitive information. Therefore, privacy of health information and diagnostic systems is indispensable.

Recent advances in the Internet of Things (IoT) and cloud computing play a vital role in the development of smart environments [8] such as smart homes and smart cities [9]. The number of senior citizens around the world will increase to 10 million in the coming decade [10]. The count of senior citizens in Japan, for example, comprises a large portion of its population, as around 21% of the population is above 65 years, a figure that will grow to 40% by 2050 [11]. Likewise, in the United States [12] and Taiwan [13], the population of people older than 65 years is 13% and 11%, respectively. According to the American Association of Retired Persons [12], 85% of senior citizens like to stay in their homes for medical assistance as long as possible. To handle the situation, smart healthcare systems can be implemented to provide efficient and cost-effective solutions [14]. Various smart healthcare systems [15, 16] have been proposed for smart homes and cities, where IoT senses the data and transmits it for evaluation at a health center. The transmitted data at the health center might be changed deliberately or accidentally (i.e., unauthorized modification), which could result in a person getting an inaccurate diagnosis. If the person is normal and the diagnosis is positive, it will cause anxiety and mental stress, which ultimately creates a negative impact on their health. On the other hand, if a person is suffering from a disease and the diagnosis is negative, it will cause complications and even the risk of death. Moreover, if the diagnosis is accurate and somehow the health information is disclosed, the patient may face adverse consequences in their social life and/or professional career.

Albeit, security and privacy are widely investigated in general [17-20], however, protection of IoT-based healthcare information is still in its infancy. Few recent studies have investigated different aspects of security and privacy in healthcare systems such as RFID [21], body area networks [22], and highlighted emerging issues and challenges [23, 24]. However, to the best of our knowledge, zero-watermarking for medical audio samples has never been employed. For example, a cloud-based healthcare framework is provided for the patients suffering from Parkinson's disease in [25], and the problem of data integrity authentication is tackled by the insertion of watermark in the speech signals of disease patient [26]. Due to embedding of watermark in speech signals, the authors reported that imperceptibility is not ideal. Nevertheless, this problem can be avoided through zero-watermarking. Zero-watermarking [27, 28] has an advantage over conventional watermarking [29] since it uses signals' features for watermarking. Unlike the traditional method, signals are not degraded as zero-

watermarking does not insert a watermark physically in a signal. In addition, the repercussions on the diagnosis accuracy after insertion of watermark and an attack on a watermarked signal is not investigated in [25]. It is very important because a privacy protected healthcare system without a trade-off between privacy and accurate diagnosis is inadequate. To avoid such predicaments, preventive controls must be implemented to ensure confidentiality and integrity of sensitive information. The main objective of this study is to design and implement a healthcare diagnostic system that protects a patient's health information, whilst also detecting vocal fold disorders by evaluating the voice sample of a subject.

Vocal fold disorders or dysphonia can be defined as an alteration in the performance and production of the voice that may interfere with communication [30]. According to the medical dictionary [31], dysphonia is a difficulty in speaking, usually evidenced by hoarseness, which represents any deviation of voice quality as perceived by the self or others [32]. Dysphonia is caused by different reasons and people suffering from it are referred to as dysphonic patients. Generally, vocal misuse including yelling, excessive talking, screaming, and crying are all irritating forces at the contact place of two vocal folds. In addition, some other factors include poor hydration, medication, alcohol consumption, and smoking [33, 34]. Vocal fold disorders can be classified into different groups depending upon the causes of occurrence. Sometimes a vocal fold disorder can occur due to abnormal growth of tissues on the vocal folds, which are benign lesions [35] that are non-cancerous in nature. Vocal fold disorders may also appear due to nerve injury that controls the vibration of the vocal folds [36], with paralysis an example of such a disorder. Another type of disorder is Keratosis, which is considered to be a pre-cancerous lesion and occurs due to the presence of unusual cells on the vocal folds [32]. Moreover, adductor is a type of spasmodic dysphonia caused by neurological disorders [37]. Detection of vocal fold disorder can be carried out through subjective assessment by using endoscopic examination of the vocal folds and different perceptual measurements [38-40]. However, possibility of human errors cannot be ignored in the subjective assessment. Therefore, various systems have been proposed for the automatic detection of the vocal fold disorders [41-45]. Such systems are useful for the self-monitoring of the patients and avoid frequent unwanted visits to consultants or practitioners. Moreover, they can also be used to provide complementary information during subjective assessment to get accurate diagnosis.

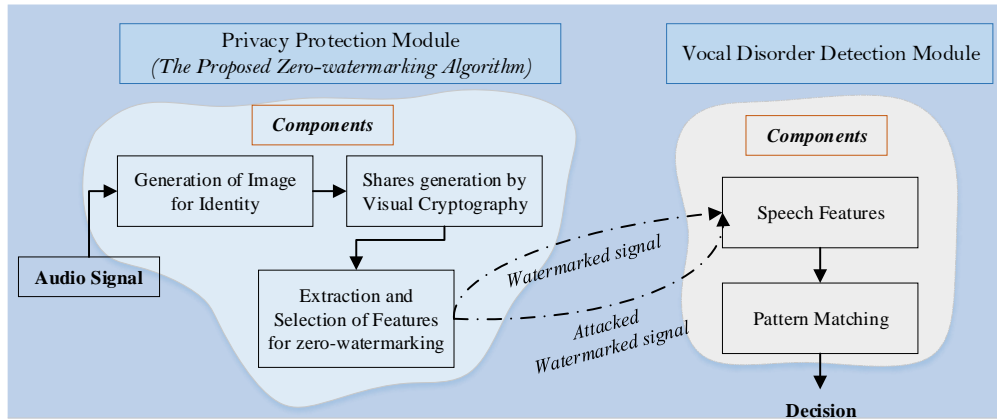
This paper presents a privacy protected healthcare system based on zero-watermarking algorithm. The proposed system mainly consists of two modules. In the first module, we propose a method for zero-watermarking to protect the identity of an individual. This is accomplished through visual cryptography by generating the two secret shares which must be available with the legitimate staff to reveal the identity of a patient. To embed identity's secret shares into the voice samples, the proposed zero-watermarking method determines the characteristics of audio samples by applying the One-Dimensional Local Binary Pattern (1D-LBP).

The second module is responsible for the detection of vocal fold disorders and is developed by using the state-of-the-art feature extraction method, Mel-frequency Cepstral Coefficients (MFCC), and a supervised machine-learning algorithm, Support Vector Machine (SVM). In addition to the baseline results for the detection of vocal fold disorders, this module also observes effect on the diagnosis accuracy due to the watermarking and noise attack.

The rest of the paper is organized as follows: Section 2 describes major components of both the modules in the proposed healthcare system. The proposed zero-watermarking algorithm, and process of embedding and extraction are explained in Section 3. Section 4 analyzes and compares the results of the proposed system with the relevant existing systems. The concluding remarks are presented in Section 5.

## 2 Privacy Protected Healthcare System

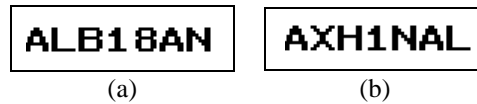
As described earlier, the proposed privacy protected healthcare system mainly features two modules i.e., privacy protection and vocal disorder detection (Figure 1). The former aims to protect privacy of an individual and it is implemented by using the proposed zero-watermarking algorithm which mainly consists of three components i.e., image generation for identity, shares generation by visual cryptography, extraction and selection of features for zero-watermarking. The vocal disorder detection system is designed for diagnosis and mainly consists of two components i.e., speech features, and pattern matching. The extracted speech features are MFCC, while, SVM is implemented for pattern matching. The following subsections describe each component of the developed system



**Figure 1:** Privacy protected healthcare system with the proposed zero-watermarking algorithm

### 2.1 Generation of Image for Identity

To generate image for subject's identity ( $S_{ID}$ ), the labels of speech signals of the voice disorder database are used. The database is recorded at the Massachusetts Eye and Ear Infirmary (MEEI) voice and speech laboratory [46]. Each audio sample in the MEEI database is labeled with an alphanumeric string of seven characters that represent a subject's identity ( $S_{ID}$ ) (Fig. 2). The last part of the alphanumeric label for dysphonic and normal subjects is AN and NAL, which indicates that the subject is normal (Fig. 2(a)) and abnormal (Fig. 2(b)) respectively. The developed healthcare system generates secret segments of the  $S_{ID}$  image for privacy protection by embedding them as a watermark.



**Figure 2:** Identity image of (a) dysphonic  $S_{ID}$  and (b) normal  $S_{ID}$  subjects.

In addition, the proposed privacy protection healthcare system is implemented and evaluated by using the MEEI database and it has been used in a number of previous studies [41-45]. The database is recorded at two different sampling frequencies, 25 KHz and 50 KHz, with a 16-bit sampling rate. The sampling frequency of all normal subjects was 50 KHz. Therefore, to be consistent, all dysphonic patients who recorded at 50 KHz are included in this study. The total number of normal audio samples is 53, and 77 dysphonic samples were selected. The list of 77 dysphonic subjects is provided in

Appendix A, in order to reproduce the results of this study. The selected subset of the MEEI database is labelled as MEEI<sub>subset</sub>, and it contains 130 audio samples. The distribution of normal and dysphonic subjects in this subset is given in Table 1.

**Table 1.** Distribution of normal and dysphonic subjects in the MEEI<sub>subset</sub>

Subjects	Gender	Number of Samples	Mean Age (Years)	Age Range (Years)	Standard Deviation (Years)
Dysphonic Patients	Male	30	48.7	18-82	18.2
	Female	47	45.8	17-79	17.9
Normal Persons	Male	21	38.8	26-59	8.5
	Female	32	34.2	22-52	7.9

## 2.2 Shares Generation by Visual Cryptography

The generated image  $S_{ID}$  of the identity is a bitonal, and therefore, it contains only black and white pixels. The black pixels are denoted by 0s and the white pixels are represented by 1s in the image of  $S_{ID}$ . In the proposed system, the secret shares of the  $S_{ID}$  image are used to protect the privacy of an individual from unauthorized access. The secret shares are generated by applying visual cryptography [47, 48]. The use of secret shares of the  $S_{ID}$  image as a watermark makes the developed healthcare system more secure than the direct usage of the  $S_{ID}$  image. The encryption process for the generation of secret shares is shown in Figure 3. The process encrypts every pixel of the  $S_{ID}$  image by replacing them with the encryption blocks  $E_1, E_2, E_3, E_4, E_5,$  and  $E_6$ , as shown in Figure 3. The corresponding encryption matrices  $I_1, I_2, I_3, I_4, I_5,$  and  $I_6$  of the blocks are:

$$I_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, I_2 = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, I_3 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, I_4 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, I_5 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, I_6 = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \quad (2)$$

The replacement of a pixel in the  $S_{ID}$  image by one of the encryption blocks means that one of the encryption matrices ( $I_1, I_2, I_3, I_4, I_5,$  and  $I_6$ ) replaces the pixel. To introduce randomness in the generated shares, a deterministic random sequence  $y$  is created by using two-dimensional Gingerbread man chaos theory, which was proposed by R. L. Devaney [49] and defined in Eq. 3.

$$\begin{aligned} x_{r+1} &= 1 - y_r + |x_r| \\ y_{r+1} &= x_r \end{aligned} \quad (3)$$

The number of encryption matrices is six, and the sequence  $y$  will select a matrix randomly to substitute a pixel in the  $S_{ID}$  image. Therefore, the values of sequence  $y$  should be in the range 1 to 6, and it is achieved by using Eq. 4.

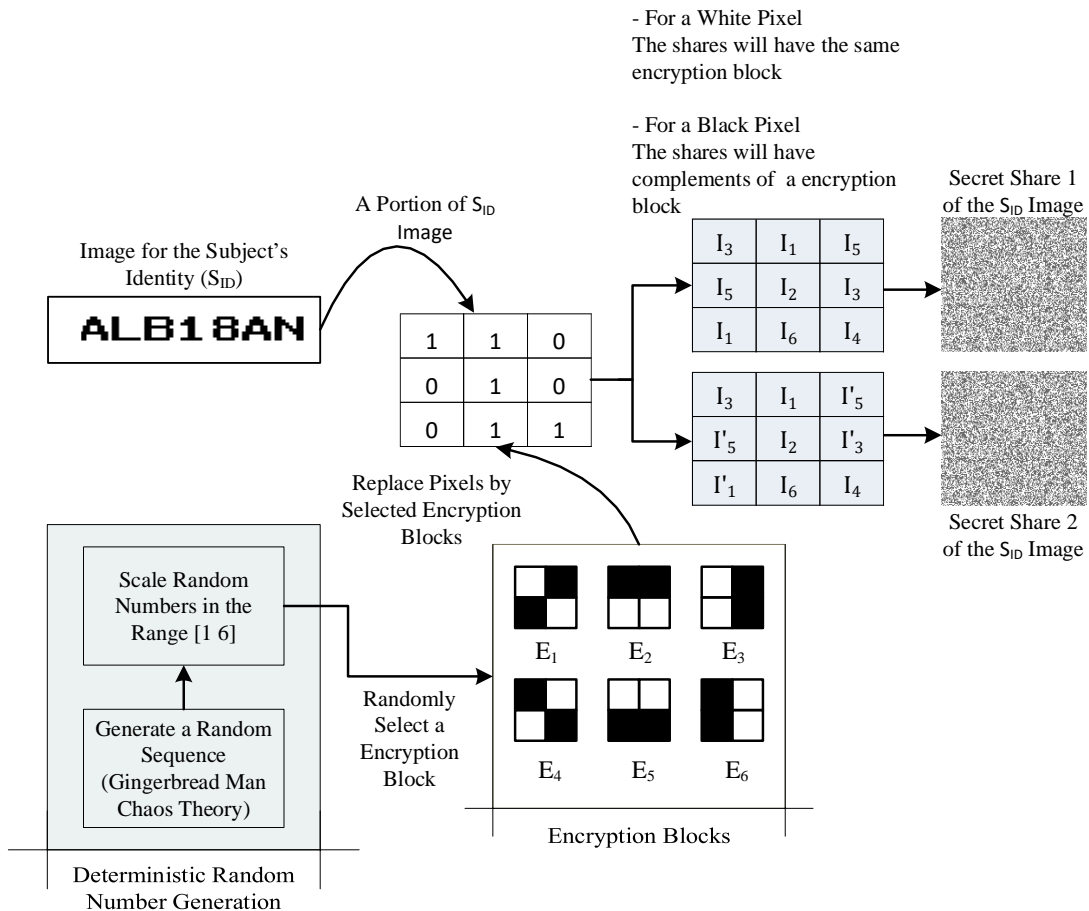
$$Z = 1 + (y \times v) \bmod 6 \quad (4)$$

where  $v$  is any number greater than 100, and division by six provides remainders in the range of 0 to 5. An encryption matrix is selected by using the values of  $Z$ , and it is substituted in the  $S_{ID}$  image according to the following criteria:

- If a pixel in the  $S_{ID}$  image is 1, then both secret shares,  $S_1$  and  $S_2$ , will have the same encryption matrix, say  $I_i$ .

- If a pixel in the  $S_{ID}$  image is 0, then both secret shares will have two different encryption matrices, which are complements of each other. Specifically, if  $I_i$  is in  $S_1$ , then  $I_i'$  is in the  $S_2$ .

For instance, if a pixel at location (1,1) in the  $S_{ID}$  image is 1, then it will be replaced by the same matrix  $I_3$  in both shares. If a pixel at location (1,3) is 0, then it will be replaced by  $I_5$  in the  $S_1$  and by  $I_5'$  in the  $S_2$ . Each pixel in the  $S_{ID}$  image is replaced by a 2x2 matrix. Therefore, the size of each secret share will be four times larger than the size of the  $S_{ID}$  image. The benefit of using 2x2 matrices is that the aspect ratio of the  $S_{ID}$  image will not be distorted because 2x2 encryption matrices double the number of rows and columns in both shares and they maintain the aspect ratio of the  $S_{ID}$  image.



**Figure 3:** Generation of secret shares for the  $S_{ID}$  image.

The insertion and extraction of the identity's secrets shares through zero-watermarking strongly dependent on the features of the audio signals. These features help in the determination of the appropriate locations for the insertion and extraction of the watermark.

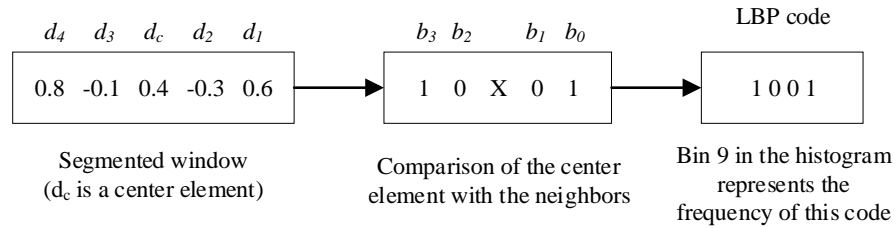
## 2.3 Extraction and Selection of Features for Zero-Watermarking

One of the most crucial steps in the proposed zero-watermarking algorithm is finding of suitable features/patterns and they can be obtained by analyzing the audio samples. Moreover, the selection of the computed features is also very vital for the reliable insertion and extraction of the identity's secret shares.

### 2.3.1 Extraction of Features

In this study, to determine the features/patterns, the audio signals are analyzed by applying the 1D-LBP operator [50, 51], which segments the audio sample into small windows and then observed the variation in the neighboring elements with respect to the center element for generation of the LBP codes. The number of codes depend on the neighbors of the center element. For  $n$  neighbors on each side of the center element, the number of LBP codes will be  $2^{2n}$ . The frequency of each LBP code can be calculated by plotting a histogram for all computed codes. Every bin of the histogram provides the information for one of the LBP codes.

In this study, to obtain 1D-LBP codes, an audio sample is segmented into the windows of five elements. The elements may be positive or negative as audio samples contain positive as well as negative amplitude. Each sample of the audio is a center element of a window. The computational steps to find the LBP code of a window of length five are shown in Figure 4. The elements of the segmented window are 0.8, -0.1, 0.4, -0.3, and 0.6. The center element is 0.4 and there are two neighbors on both sides. The number of LBP codes in this case is  $2^4$  in the range from 0 to  $2^4-1$ . These codes were obtained by comparing the center element with its neighbors. If a neighbor is equal to or greater than a center element, then it is replaced by a 1 in the window. However, if a neighbor is smaller than the center elements, then it is replaced by a 0. In Figure 4, the neighbors 0.8 and 0.6 are greater than the center element 0.4, hence they are replaced by 1s, whereas -0.1 and -0.3 are replaced by 0s because they are smaller than the center element. As a result, a 4-bit binary number  $(1001)_2$  is obtained, where the bit at the extreme left is the most significant bit. The frequency of this code is represented by the corresponding bin, which is 9, since  $(9)_{10}=(1001)_2$ .



**Figure 4:** Computational step for 1D-LBP codes.

The 1D-LBP code for a window  $w$  with the center element  $d_c$  and the neighbors  $d_1$ ,  $d_2$ ,  $d_3$ , and  $d_4$  can be calculated using Eq. 1 as follows:

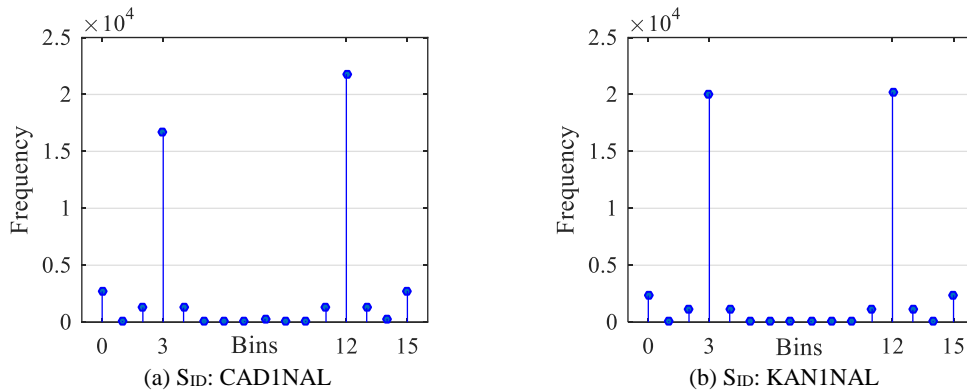
$$b_{j-1} = \begin{cases} 1 & \text{if } d_j \geq d_c \\ 0 & \text{if } d_j < d_c \end{cases} \quad \text{where } j=1, 2, 3, 4 \quad (1)$$

where  $d_1$  and  $d_2$  are two neighbors on the right side of the center element,  $d_3$  and  $d_4$  are two neighbors on the left side,  $b_3b_2b_1b_0$  is a required 1D-LBP code, and  $b_3$  is the most significant bit.

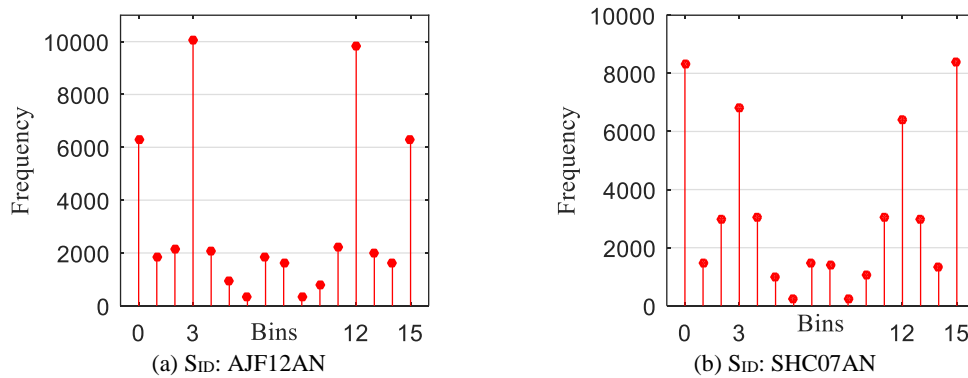
### 2.3.2 Selection of Features

The selection of locations for zero-watermarking are determined by analyzing the histograms of the computed 1D-LBP codes. The histograms representing the 1D-LBP codes for two different normal subjects are plotted in Figures 5(a) and 5(b). Moreover, the histograms for two different dysphonic subjects are shown in 6(a) and 6(b). It can be observed that LBP codes 0000 in bin 0, 0011 in bin 3, 1100 in bin 12, and 1111 in bin 15 have high frequencies in the histograms of normal subjects. For instance, in the histogram of CAD1NAL, the frequency of these four LBP codes is 2706, 16663, 21785, and 2747, respectively. The sum frequency of these codes is 43901, which is 90% of the total number of codes in all bins. These codes provide sufficient capacity for zero-watermarking. Codes 0000 and 1111 represent that there are no variations in neighboring elements with respect to the center elements. Code 0000 means that all neighbors are smaller than center elements, and code 1111 represents that all neighbors are greater than the center. Code 1100 shows that the neighbors on the left are greater, and on the right they are smaller than the center elements.

In addition, the same kind of trend is found in the histograms of dysphonic subjects. Similar to the histograms of normal subjects, LBP codes 0000, 0011, 1100, and 1111 also have high frequencies in dysphonic subjects. In Figure 6, the frequencies of these codes in the histogram of AJF12AN are 6299, 10028, 9833, and 6307, respectively. The sum frequency of these codes is 32467, which is 65% of the total number of codes. These codes provide enough capacity for zero-watermarking.



**Figure 5:** Histograms of 1D-LBP codes for normal subjects (a) CAD1NAL (b) KAN1NAL



**Figure 6:** Histograms of 1D-LBP codes for dysphonic subjects (a) AJF12AN (b) SHC07AN



## 2.4 Speech Features

The first component of the second module is extraction of MFCC features. The MFCC is a well-known speech feature and has been used successfully in a number of studies for the detection of vocal fold disorders [52-54]. In this study, the MFCC [55] is used to investigate the effect on diagnosis of the vocal fold disorders due to the embedding and extraction processes of watermarks and noise attacks on the watermarked audio. The MFCC mimics the human hearing system and behaves like an expert clinician in diagnosing voice disorders. Prior to the computation of MFCC, it is important to divide an audio into short frames. Moreover, each frame is multiplied with a hamming window to ensure the continuity of divided frames into successive frames, and to avoid the spectral leakage at the ends of segmented frames during implementation of Fourier's transformation (FT). The FT converts the audio from the time domain to the frequency domain, and the result of this step is referred to as the spectrum of the audio. Then, the spectrum is filtered through a Mel-spaced band-pass filter bank. The bandwidth of each filter is called a critical bandwidth, which is one of the principles of human psychoacoustic principles [56]. The Mel scale is given by Eq. 5, and it is linear up to 800 Hz and logarithmic beyond that. In Eq. 5,  $h$  represents the frequency in Hz and  $m$  stands for the corresponding frequency in Mel scale.

$$m = 2595 \log_{10} \left( 1 + \frac{h}{700} \right) \quad (5)$$

The logarithm is applied to the Mel-scaled spectrum for the compression, and it converts the multiplication operations into additives to make the calculation easier. The last step is the application of discrete cosine transformation to decorrelate the coefficients. The output of the last step is the required MFCC features. The MFCC are calculated for all speech signals of MEEI<sub>subset</sub>. With the combination of MFCC and SVM, the baseline results for the detection of disorder will be obtained. Ultimately, these results will be compared with those obtained after watermarking and noise attacks.

## 2.5 Pattern Matching

The extracted MFCC features are given to SVM for differentiation of normal and disordered signals. SVM is a supervised learning algorithm of machine learning, which is considered to be one of the most successful approaches in pattern recognition [57, 58]. SVM learns patterns from the given training data and uses them to predict the label for an object. In this study, SVM is implemented to distinguish between normal and dysphonic subjects. The ultimate goal of the SVM is to determine an optimal hyperplane that provides maximum distances between the samples of normal and dysphonic subjects. SVM is implemented with linear and radial basis function (RBF) kernels. In most of the cases, the classes are not linearly separable. Therefore, an RBF kernel is also used, which maps the original input space to a higher dimensional space to separate the classes. The RBF kernel is given by Eq. 6, where  $x$  and  $x'$  represent the training and testing data, respectively, and  $\gamma$  is a free parameter.

$$F(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (6)$$

## 3 The Proposed Zero-Watermarking Algorithm

In the developed healthcare system, the privacy protection module is built by proposing and implementing a new zero-watermarking algorithm. The embedding and extraction processes of the proposed algorithm describe the steps for the insertion and recovery of the identity.

### 3.1 Embedding Process for the Insertion of Identity

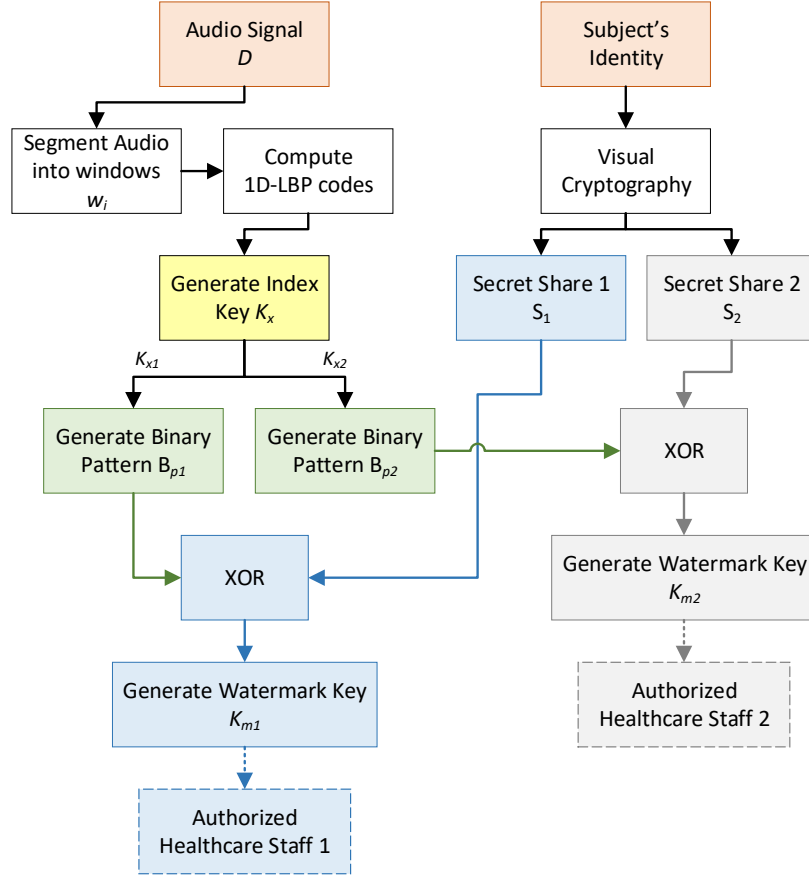
The embedding process of the proposed algorithm to insert the identity of a subject is depicted in Figure 7. The steps for the embedding process are described as follows:

1. First, the proposed algorithm generates an image for  $S_{ID}$ , as shown in Figure 2, with dimension  $a \times b$ .
2. In the next step, the algorithm creates two secret shares,  $S_1$  and  $S_2$ , of the  $S_{ID}$  image by using the visual cryptography explained in Section 2.3. The dimensions of each secret share are  $2a \times 2b$ , i.e.,  $dimS_1 = 2a \times 2b$  and  $dimS_2 = 2a \times 2b$ . The desired watermarks are secret shares  $S_1$  and  $S_2$ , given by Eq. 7.

$$S_1 = \begin{bmatrix} s_{(1,1)}^1 & s_{(1,2)}^1 & s_{(1,3)}^1 & \cdots & s_{(1,2b)}^1 \\ s_{(2,1)}^1 & s_{(2,2)}^1 & s_{(2,3)}^1 & \cdots & s_{(2,2b)}^1 \\ s_{(3,1)}^1 & s_{(3,2)}^1 & s_{(3,3)}^1 & \cdots & s_{(3,2b)}^1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{(2a,1)}^1 & s_{(2a,2)}^1 & s_{(2a,3)}^1 & \cdots & s_{(2a,2b)}^1 \end{bmatrix} \quad S_2 = \begin{bmatrix} s_{(1,1)}^2 & s_{(1,2)}^2 & s_{(1,3)}^2 & \cdots & s_{(1,2b)}^2 \\ s_{(2,1)}^2 & s_{(2,2)}^2 & s_{(2,3)}^2 & \cdots & s_{(2,2b)}^2 \\ s_{(3,1)}^2 & s_{(3,2)}^2 & s_{(3,3)}^2 & \cdots & s_{(3,2b)}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{(2a,1)}^2 & s_{(2a,2)}^2 & s_{(2a,3)}^2 & \cdots & s_{(2a,2b)}^2 \end{bmatrix} \quad (7)$$

3. The algorithm segments the host audio signal  $D$  into the windows  $w_i$  of size  $2n+1$ , as mentioned in Eq. 8, so that each element of the audio  $D$  is a center element of a window.

$$D = [w_1, w_2, w_3, \dots, w_g]^T \text{ and } 1 \leq g \leq \left\lfloor \frac{N-2n}{2n+1} \right\rfloor \quad (8)$$



**Figure 7:** Embedding process of the proposed algorithm for insertion of  $S_{ID}$ .

In Eq. 8,  $N$  is the number of samples in the host audio signal  $D$ ,  $n$  is the number of elements on each side of the center element in a window, and  $\lfloor \cdot \rfloor$  is a floor operator. In this study, we use  $n=2$ .

4. The next step is to compute the 1D-LBP code for each window  $w_i$  by using Eq. 1, and then to identify the windows whose 1D-LBP codes are 0000, 0011, 1100, or 1111, and store indices of such windows in an index key  $K_x$  as:

$$K_x^k = \{i \mid LBP(w_i) \in [0011, 1111, 0000, 1100], k = 1, 2, 3, \dots, \dim S_1 + \dim S_2\} \quad (9)$$

where  $LBP(w_i)$  denotes the 1D-LBP code for a window  $w_i$ .

5. The next step is to partition the index key  $K_x$  into two equal parts, say  $K_{x1}$  and  $K_{x2}$ . After this, keys  $K_{x1}$  and  $K_{x2}$  should be reshaped according to the dimension of the secret shares, which is  $2a \times 2b$ . Use the indices of  $K_{x1}$  and  $K_{x2}$  and compute 1D-LBP codes for the respective windows  $w_i$  to generate a binary pattern,  $B_{p1}$  and  $B_{p2}$ , with the following criteria:

- The elements of  $B_{p1}$  and  $B_{p2}$  will be 1 if the bits on the right side of the center element are greater than the center element and both bits of the left side are either greater or less, i.e., 0011 and 1111.
- The elements of  $B_{p1}$  and  $B_{p2}$  will be zero if the bits on the right side are less than the center element and both bits on the left side are greater or less, i.e., 0000 and 1100. The criteria is summarized in Eq. 10 and 11 as:

$$B_{p1}(e, f) = \begin{cases} 1 & \text{if } LBP(w_{K_{x1}^{(e,f)}}) \in [0011, 1111] \\ 0 & \text{if } LBP(w_{K_{x1}^{(e,f)}}) \in [0000, 1100] \end{cases} \quad (10)$$

$$B_{p2}(e, f) = \begin{cases} 1 & \text{if } LBP(w_{K_{x2}^{(e,f)}}) \in [0011, 1111] \\ 0 & \text{if } LBP(w_{K_{x2}^{(e,f)}}) \in [0000, 1100] \end{cases} \quad (11)$$

where  $e=1,2,3, \dots, 2a$  and  $f=1,2,3, \dots, 2b$ .

6. The watermark detection key  $K_{m1}$  for secret share  $S_1$  is obtained by performing exclusive OR (XOR) operation between the generated binary pattern  $B_{p1}$  and the  $S_1$  as given in Eq. 12.

$$K_{m1} = B_{p1} \oplus S_1 \quad (12)$$

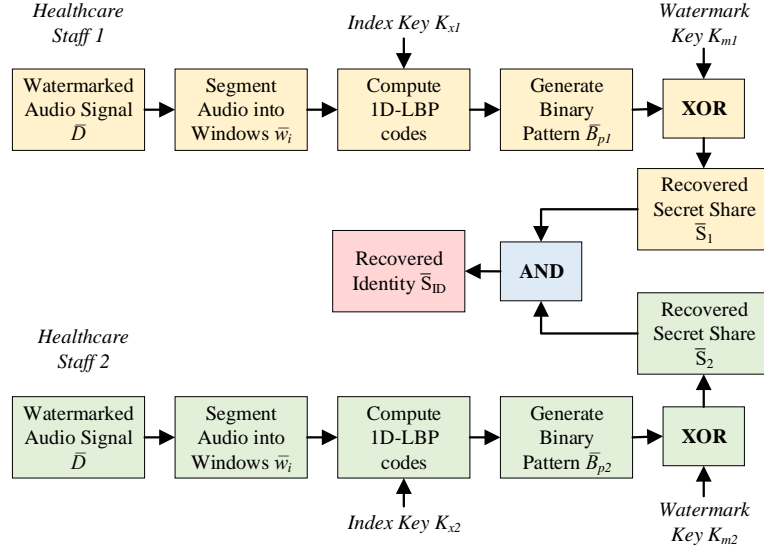
7. Similarly, the watermark detection key  $K_{m2}$  for secret share  $S_2$  is obtained by performing XOR operation between the generated binary pattern  $B_{p2}$  and the  $S_2$  as given in Eq. 13.

$$K_{m2} = B_{p2} \oplus S_2 \quad (13)$$

Finally, the host audio signal  $D$  with two keys, the index key  $K_{x1}$ , and the watermarking detection key  $K_{m1}$  will be transmitted to the authorized healthcare staff 1. Likewise, the host audio sample  $D$  with  $K_{x2}$  and  $K_{m2}$  will be sent to the authorized healthcare staff 2. The purpose is to ensure collusion and split knowledge i.e., separation of duties.

### 3.2 Extraction Process for the Recovery of Identity

Two healthcare staff have the keys and transmitted audios. The block diagram for the extraction process of the proposed zero-watermarking algorithm is shown in Fig. 8. The steps of the extraction process to recover the identity are:



**Figure 8:** Extraction process of the proposed algorithm to recover the identity  $S_{ID}$ .

1. The algorithm segments the watermarked audio signal  $\bar{D}$  into the windows  $\bar{w}_i$ , as shown in Eq. 14, of size  $2n+1$  so that each element of  $\bar{D}$  is a center element of a window.

$$\bar{D} = [\bar{w}_1, \bar{w}_2, \bar{w}_3, \dots, \bar{w}_g]^T \text{ and } 1 \leq g \leq \left\lfloor \frac{N-2n}{2n+1} \right\rfloor \quad (14)$$

In Eq. 14,  $N$  is the number of samples in the test audio signal  $\bar{D}$ ,  $n$  is the number of elements on each side of the center element in a window, and  $\lfloor \cdot \rfloor$  is a floor operator.

2. In the next step, calculate the 1D-LBP codes for the segmented windows whose indices are listed in the index key  $K_{x1}$ . Then, estimate the binary pattern  $\bar{B}_{p1}$  by using the criteria given in Eq. 15:

$$\bar{B}_{p1}(e, f) = \begin{cases} 1 & \text{if } LBP(\bar{w}_{K_{x1}^{(e,f)}}) \in [0011, 1111] \\ 0 & \text{if } LBP(\bar{w}_{K_{x1}^{(e,f)}}) \in [0000, 1100] \end{cases} \quad (15)$$

where  $e=1,2,3, \dots, 2a$  and  $f=1,2,3, \dots, 2b$ .

3. After that, XOR operation between the estimated pattern  $\bar{B}_{p1}$  and the watermark detection key  $K_{m1}$  is performed to recover the secret share  $\bar{S}_1$  of the  $S_{ID}$  image as:

$$\bar{S}_1 = \bar{B}_{p1} \oplus K_{m1} \quad (16)$$

4. Likewise, use steps 1 to 3 to recover the secret share  $\bar{S}_2$  of the  $S_{ID}$  image by using the index key  $K_{x2}$  and the watermark key  $K_{m2}$ .
5. The benefit of using the visual cryptography is that no further calculation is required to decrypt the recovered secret share  $\bar{S}_1$  and  $\bar{S}_2$  to disclose the identity of a subject. To recover the identity, print the  $\bar{S}_1$  and  $\bar{S}_2$  on two different transparencies, and superimpose them on each other.

Alternatively, perform AND operation between the recovered secret shares  $\bar{S}_1$  and  $\bar{S}_2$  as given in Eq. 17.

$$\bar{S}_{ID} = \bar{S}_1 \wedge \bar{S}_2 \quad (17)$$

where  $\bar{S}_{ID}$  is the recovered identity of a subject.

## 4 Experimental Results and Discussion

The second module of the developed healthcare system for vocal disorder detection is implemented with MFCC and SVM. This module contains two important phases i.e., training and testing. The former takes labeled audio samples and extracts MFCC features. For the computed features, SVM generates the model for each type of subject. In SVM, dysphonic subjects are designated as a positive class, and normal subjects are specified as a negative class. The testing phase takes unlabeled/unknown audio samples and calculates the MFCC features. Then, SVM uses these features to predict the class of unknown audio samples through pattern matching. In both phases, the MFCC features are calculated by using a frame size of 512 samples, a hamming window with 512 points to taper the ends of the divided frames, and 29 band-pass filters are used in a Mel-spaced filter bank.

Moreover, the  $k$ -fold cross validation approach is applied to obtain the results for disorder detection. In this way, bias of the training and testing audio samples can be avoided. In this study, the 3-fold cross-validation approach is used to divide all audio samples into three disjointed subsets. Each time, the system is tested with one of the sets, while it is trained with the remaining two sets of audio samples. The performance of the proposed system is gauged based on the following metrics: sensitivity (SEN), specificity (SPE), and accuracy (ACC). These metrics are defined as:

- Sensitivity (SEN): A ratio between accurately detected audio samples of the dysphonic subjects by the system and the total number of dysphonic subjects.
- Specificity (SPE): A ratio between correctly detected audio recordings of the normal subjects by the system and the total number of normal subjects.
- Accuracy (ACC): A ratio between the total number of truly detected samples by the system and the total number of audio samples.

In addition, the area under the Receiver Operating Characteristic (ROC) is also used as a performance metric. The area under the ROC curve (AUC) shows the reliability of the developed system in diagnosing disorders.

The detection results of the developed healthcare system for the  $MEEI_{\text{subset}}$  are provided in Table 2. The ACC of the system with an RBF kernel is better than the ACC obtained with a linear kernel, which suggests that data are not linearly separable, and the RBF mapped the original space into a higher dimension space to obtain the optimized hyperplane for the normal and dysphonic subjects. Moreover, the AUC of the RBF is greater than the AUC of the linear kernel, which shows that the RBF kernel is more reliable in diagnosing voice disorders.

**Table 2.** Results of Disorder Detection

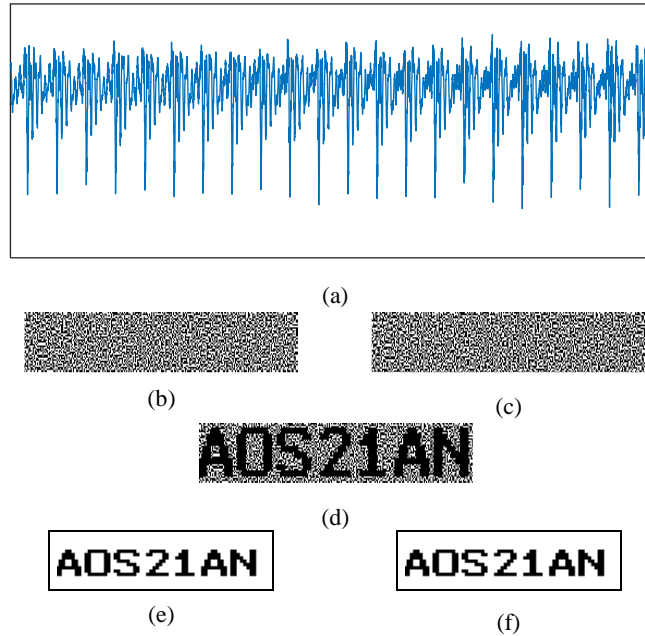
Kernel	% SEN $\pm$ STD	% SPE $\pm$ STD	% ACC $\pm$ STD	AUC
Linear	96.32 $\pm$ 4.1	81.96 $\pm$ 6.7	90.64 $\pm$ 5.8	0.89

RBF	98.72%±2.2	83.22%±8.3	92.39%±4.4	0.95
-----	------------	------------	------------	------

In the following subsections, the proposed algorithm is evaluated through a performance test of the algorithm in identity recovery, imperceptibility of embedded identity, detection reliability to recover the identity, and, lastly, robustness of the algorithm against the noise of various signal-to-noise ratios (SNR). In each case, we also observe the effect on the ACC for disorder detection in the developed healthcare system. Whilst our prime concern is privacy, this should not be at the cost of inaccurate diagnosis.

#### 4.1 Performance Test

The audio signal with the inserted and recovered identities are depicted in Fig. 9. An audio sample of a subject is taken from the MEEI<sub>subset</sub> database, and is shown in Figure 9(a). The privacy protection module of the proposed system generates two secret shares,  $S_1$  and  $S_2$ , for the subject's identity in order to conceal it. The image for the identity of subject AOS21AN is shown in Figure 9(e), and the two generated secret shares,  $S_1$  and  $S_2$ , are depicted in Figures 9(b) and 9(c). The dimension of the  $S_{ID}$  image for all audio samples in the MEEI<sub>subset</sub> is 20 x 126. After generating the secret shares, the embedding process of the proposed zero-watermarking algorithm created the index keys  $K_{x1}$ ,  $K_{x2}$  and the watermark key  $K_m$  for insertion of identity. To disclose the identity of a subject, it is compulsory to have access to both shares at the same time.



**Figure 9:** The insertion and recovery of the identity  $S_{ID}$  in an audio signal (a) An original audio sample, (b) the first secret share of the identity  $S_1$ , (c) the second secret share of the identity  $S_2$ , (d) the recovered identity  $\bar{S}_{ID}$ , (e) the original image for identity  $S_{ID}$ , and (f) the recovered identity image  $\hat{S}_{ID}$  with the procedure given in Eq. 18.

The extraction process of the privacy protection module in the proposed algorithm uses the keys  $K_{x1}$ ,  $K_{x2}$ , and  $K_m$ , along with the transmitted audio sample to extract the secret share by using the patterns (1D-LBP code) of the audio samples. The extracted secret shares,  $\bar{S}_1$  and  $\bar{S}_2$ , will be printed on the transparencies, and their superimposing on each other will disclose the identity of the patients.

The disclosed identity of the patient is shown in Figure 9(d). The background of the recovered identity is not white as in the original image of the identity (Figure 9(e)), which is due to the visual cryptography. Visual cryptography does not require any computation for decryption to disclose the identity. However, some computational steps are required to retrieve exactly the same image of identity.

To obtain exactly the same identity, a simple procedure containing some additional steps is implemented in the proposed zero-watermarking algorithm. This procedure takes both recovered secret shares,  $\bar{S}_1$  and  $\bar{S}_2$ , as input and compares the corresponding blocks of  $2 \times 2$  dimensions in both shares. If the corresponding  $2 \times 2$  blocks in both shares are the same, then a pixel is 1 in the recovered identity. In case of different corresponding blocks, the pixel in the recovered identity is 0. The recovered identity used in this procedure is depicted in Figure 9(f), and it is similar to the original identity image shown in Figure 9(e). The procedure is defined in Eq. 18 as:

$$\tilde{S}_{ID}(a', b') = \begin{cases} 1 & \text{if } s_{b1}(u', v') = s_{b2}(u', v') \\ 0 & \text{if } s_{b1}(u', v') \neq s_{b2}(u', v') \end{cases}$$

where (18)

$$u' = 2a' - 1 : 2a'$$

$$v' = 2b' - 1 : 2b'$$

$$a' = 1, 2, 3, \dots, 2a \quad \text{and} \quad b' = 1, 2, 3, \dots, 2b$$

where colons ‘:’ are used to represent the range,  $s_{b1}$  and  $s_{b2}$  are  $2 \times 2$  blocks of  $\bar{S}_1$  and  $\bar{S}_2$ , respectively. We have implemented this procedure because we want to demonstrate the objective evaluation of the proposed zero-watermarking algorithm. The objective evaluation is only possible when the recovered identity is of the same size and background. For an objective evaluation, we used Normalized Cross-Correlation (NCR), Bit Error Rate (BER), and Energy Ratio (ENR). These performance metrics are defined in equations 19, 20 and 21. In Eq. 20,  $t$  represents the number of erroneously extracted bits.

$$\text{NCR}(S_{ID}, \tilde{S}_{ID}) = \frac{\sum_{i=1}^a \sum_{j=1}^b S_{ID}(i, j) \tilde{S}_{ID}(i, j)}{\left( \sum_{i=1}^a \sum_{j=1}^b S_{ID}^2(i, j) \right)^{\frac{1}{2}} \left( \sum_{i=1}^a \sum_{j=1}^b \tilde{S}_{ID}^2(i, j) \right)^{\frac{1}{2}}} \quad (19)$$

$$\text{BER}(\%) = \frac{t}{a \times b} \times 100 \quad (20)$$

$$\text{ENR}(S_{ID}, \tilde{S}_{ID}) = \frac{\sum_{i=1}^a \sum_{j=1}^b \tilde{S}_{ID}^2(i, j)}{\sum_{i=1}^a \sum_{j=1}^b S_{ID}^2(i, j)} \quad (21)$$

The metrics NCR, BER, and ENR for the audio sample of patient AOS21AN are 1, 0, and 1, respectively. The NCR equals to 1 describes that the original image of identity  $S_{ID}$  and recovered image of identity  $\tilde{S}_{ID}$  are identical, and BER equals to 0 refers to the fact that there is no difference between the pixels of both images. Moreover, the ENR equals to 1 represents that the energy of the



original and the recovered identity images is the same. These metrics are computed for all audio samples of the MEEI<sub>subset</sub> database and listed in Table 3.

**Table 3:** Performance of the proposed algorithm for the MEEI<sub>subset</sub>

Modules	Performance Metrics			
Privacy protection	NCR: 1	BER: 0	ENR: 1	
Vocal disorder detection	SEN: 98.72%±2.2	SPE: 83.22%±8.3	ACC:92.39%±4.4	AUC:0.95

The experimental results (NCR=1, BER=0 and ENR=1) demonstrate that the performance of the proposed algorithm is excellent. The values of NCR, BER, and ENR show that there is no difference between the original and recovered identity. In addition, it can also be seen in Table 3 that the SEN, SPE, ACC, and AUC for disorder detection are not affected, and they are similar to the baseline result provided in Table 2. We have used only the RBF kernel in these experiments since this kernel also performs better for the baseline results.

## 4.2 Imperceptibility

In watermarking algorithms, one of the major problems is the degradation of the audio quality due to embedded watermarks. Inaudibility of the inserted watermark demonstrates the success of an algorithm. Imperceptibility in the privacy protected healthcare diagnostic system is very crucial. A significant difference between the host and watermarked audio samples will lead to a false diagnosis. In this study, the proposed zero-watermarking algorithm did not insert the identity of a subject tangibly into the audio sample. In fact, the identity was inserted into the secret key. Therefore, the watermarked audio  $\bar{D}$  and the host audio  $D$  are exactly same and there is no chance for the audibility of the watermark. It is a positive aspect of the proposed zero-watermarking algorithm that the accuracy of diagnosis in the healthcare system will not be affected.

The objective analysis of imperceptibility is done by computing the SNR of host audio  $D$  and watermarked audio  $\bar{D}$ . The SNR was calculated for all audio samples in the MEEI<sub>subset</sub> database and their corresponding watermarked audio samples by using the relation provided in Eq. 22. For all audio samples, the  $SNR = \infty$  demonstrates that the proposed algorithm does not affect audio samples during watermark embedding and extraction processes. Moreover, the performance of Module 2 for disorder detection is also not affected and the results are provided in Table 4. The results of SEN, SPE, ACC, and AUC are similar to the baseline results.

**Table 4:** Performance of the proposed algorithm for the MEEI<sub>subset</sub> in case of imperceptibility

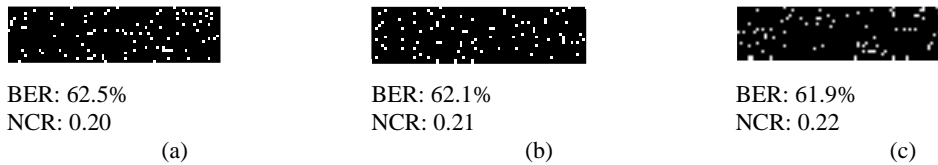
Modules	Performance Parameters			
Privacy protection	SNR = $\infty$			
Vocal Disorder Detection	SEN: 98.72%±2.2	SPE: 83.22%±8.3	ACC:92.39%±4.4	AUC:0.95

$$SNR(D, \bar{D}) = 10 \log_{10} \left( \frac{\sum_{i=1}^N d_i^2}{\sum_{i=1}^N (d_i - \bar{d}_i)^2} \right) \quad (22)$$

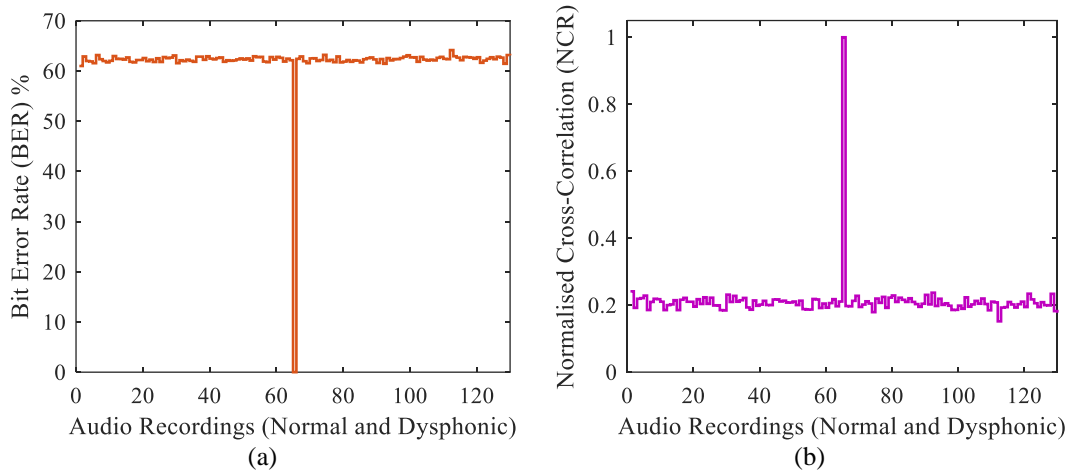
### 4.3 Detection Reliability

Detection reliability investigates whether the proposed zero-watermarking algorithm has the undesired property of watermark extraction by using secret keys of a different subject. For instance, consider an audio sample of a subject with identity KLC06AN. The proposed algorithm embedded the subject's identity, and generated the index and watermark keys  $K_{x1}$ ,  $K_{x2}$ , and  $K_m$ . To examine the detection reliability of the proposed algorithm, we examined whether it was possible to extract the identity from the audio samples of subjects KAN1NAL, KAS09AN, and KAH02AN by using the keys of subject KLC06AN.

The identities extracted from samples KAN1NAL, KAS09AN, and KAH02AN are shown in Figures 10(a), 10(b), and 10(c), respectively. The identities of these subjects are not revealed by using the secret keys of subject KLC06AN, which confirms the detection reliability of the proposed algorithm.



**Figure 10:** Attempt to extract the identities by using the keys of KLC06AN from audio samples (a) KAN1NAL (b) KAS09AN (c) KAH02AN



**Figure 11:** Detection reliability of the proposed zero-watermarking algorithm by using (a) BER and (b) NCR

Moreover, by utilizing the secret key of subject KLC06AN, the identities are extracted from all samples of the  $MEEI_{\text{subset}}$ . The extracted identities are compared with the embedded identity of the subject KLC06AH by computing BER and NCR. The values of BER and NCR for all 130 audio samples are depicted in Figures 11(a) and 11(b), respectively. Figure 11(a) shows that the BER for extracted identities of all audio samples is above 60%, which means that identities can be revealed only when relevant keys are available. Figures 10(a)-(c) confirm that extracted identities with a BER above 60% do not provide any information about the identities. These extracted identities are just a random collection of pixels. However, the BER is zero for one sample, which suggests that the secret


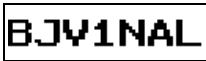

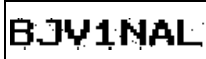






keys belong to that audio sample and the identity of that subject is KLC06AN. Furthermore, in Figure 11(b), the NCR for all extracted identities is below 0.25, which suggests that extracted identities with the keys of a different subject and the embedded identity are entirely different. It can be noticed that the NCR is 1 for one sample, which signifies that the keys belong to that audio sample. Therefore, it can be concluded that the proposed algorithm detects the identity of a subject reliably, and a key of a different subject cannot be used to disclose the identity of some other subject.

#### 4.4 Robustness

A watermark algorithm should be robust against malicious attacks. One of the most commonly used attacks involves the addition of noise to distort the watermarked audio in order to eradicate the identity of a subject. In the healthcare diagnosis system, a trade-off between the privacy and accurate diagnosis of an individual is very crucial and of extreme importance. In this study, the White-Gaussian noise of various SNR is added to the watermarked audio to examine the robustness of the proposed algorithm.

Table 5 provides the results of the proposed algorithm after adding the noise in the watermarked audio sample. In case of 60dB SNR, the identity extracted from the attacked audio sample  $\hat{S}_{ID}$  and original identity  $S_{ID}$  has only 1% BER, whereas the other metrics (NCR and ENR) are very good. Moreover, it can be seen that the recovered identity is also not distorted. The BER is increased to 3% for 50 dB and 10% for 40 dB. However, the recovered identity  $\hat{S}_{ID}$  is recognizable in both situations. Furthermore, a significant distortion is noted in the recovered identity  $\hat{S}_{ID}$  for an attack of 30 dB, but at the same time, due to the attack of high SNR, the diagnostic system of vocal disorder detection module failed to diagnose the attacked audio accurately.

**Table 5:** The performance of the proposed zero-watermarking algorithm for noise attack

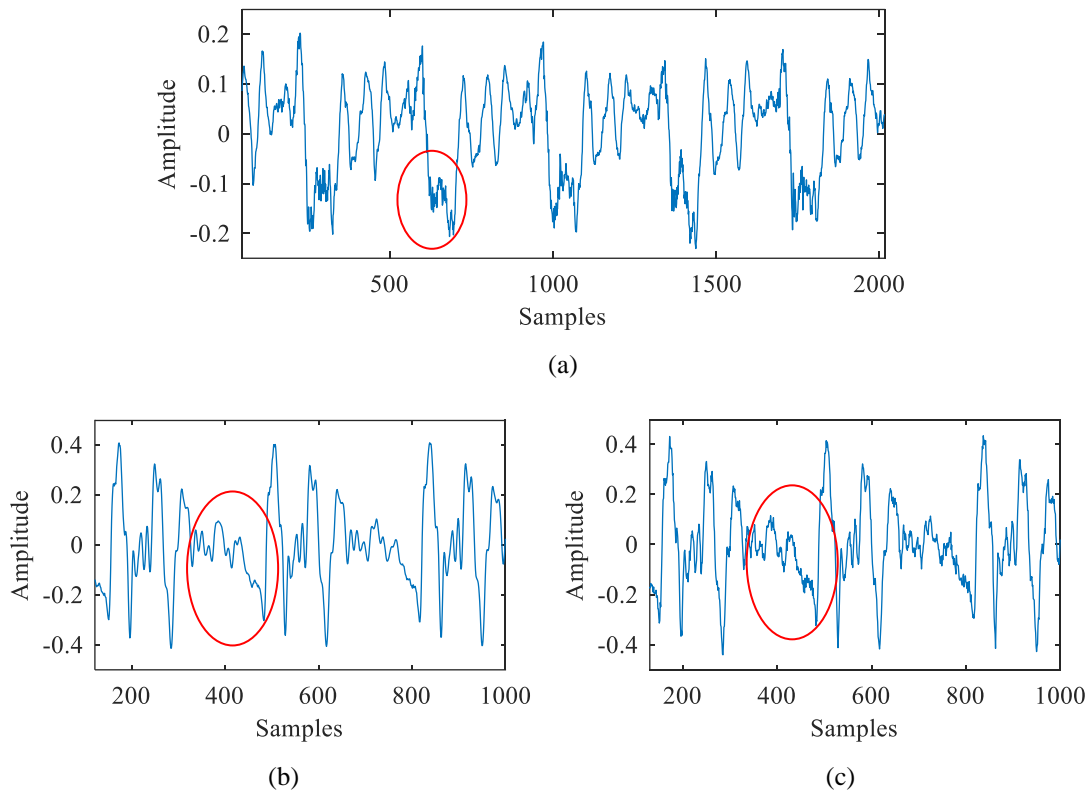
dB	Recovered Identity of the Subject		NCR	BER	ENR	Diagnosis
	$\bar{S}_{ID}$	$\hat{S}_{ID}$				
No			1	0	1	True
60			0.99	1.11	0.98	True
50			0.98	3.01	0.95	True
40			0.92	9.64	0.85	True
30			0.81	22.61	0.65	False

In Table 6, the detection results of vocal disorder detection module for the diagnosis of voice disorders are listed for noise attacks of 60 dB, 50 dB, 40 dB, 30 dB, and 20 dB. At 60 dB, the SEN is 98.7%, SPE is 83.2%, and ACC is 92.4%. Table 6 shows that SPE decreases as SNR increases. The reason is that after increasing the SNR, the audio sample of a normal sample becomes transient and complex, similar to that of a dysphonic subject, as shown in Figure 12. Due to voice disorders, the vocal folds of a dysphonic patient vibrate irregularly and generate complex patterns in the audio signal. This makes the voice of a dysphonic subject sound noisy, harsh, and strained to the ears. Therefore, the diagnosis system misclassified a normal subject as a dysphonic one after a noise attack

of high SNR, which affects the SPE of the system. The misclassification of normal subjects ultimately affects the ACC of the developed healthcare system. Furthermore, attacks of 30 dB and 20 dB are audible, and so healthcare staff can ask the subject to transmit another audio sample for the diagnosis.

**Table 6:** The results of the developed healthcare system for noise attacks of different SNR

dB	% SEN $\pm$ STD	% SPE $\pm$ STD	% ACC $\pm$ STD	AUC
No	98.72% $\pm$ 2.2	83.22% $\pm$ 8.3	92.39% $\pm$ 4.4	0.95
60	98.72% $\pm$ 2.2	83.22% $\pm$ 8.3	92.39% $\pm$ 4.4	0.95
50	98.72 $\pm$ 2.22	81.37 $\pm$ 11.3	91.63 $\pm$ 6.2	0.89
40	98.71 $\pm$ 2.22	77.45 $\pm$ 9.3	90.08 $\pm$ 5.4	0.88
0	100 $\pm$ 0	33.99 $\pm$ 6.2	73.15 $\pm$ 3.1	0.60
20	100 $\pm$ 0	0 $\pm$ 0	59.23 $\pm$ 1.2	0.31



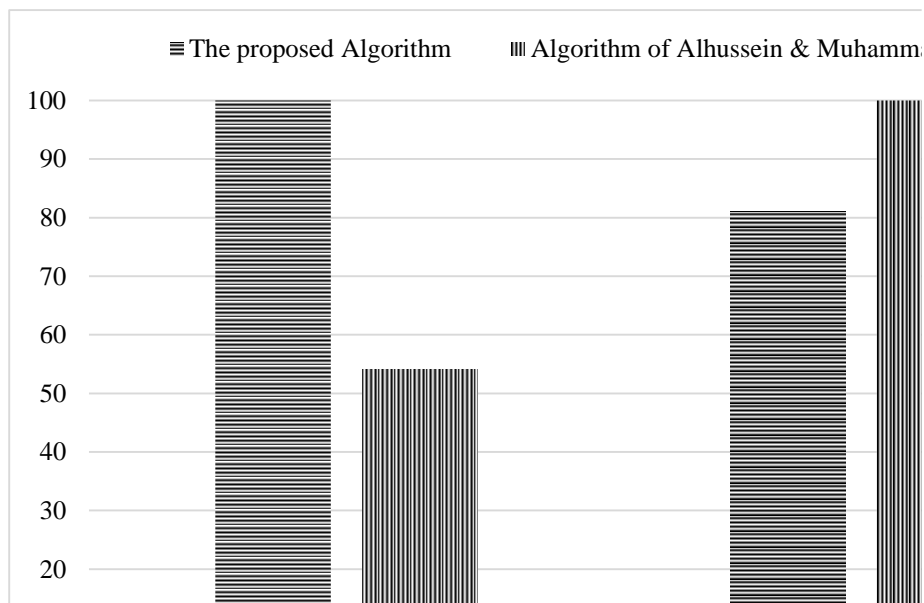
**Figure 12:** (a) Dysphonic Subject, (b) Normal subject, (c) Normal subject after adding White-Gaussian noise of 30 dB

## 4.5 Comparison

Despite our best efforts, we did not find any privacy protected healthcare system implemented with zero-watermarking. However, a framework of a healthcare system for Parkinson's disease is

presented in [25], in which the image of a subject’s identity is embedded by using a traditional watermarking algorithm in the speech signal of an affected subject. The authors concluded that the watermark algorithm has a high NCR between the original and recovered identity, but at the same time the SNR is not good. The reason is that the traditional watermarking algorithms insert watermarks in speech signals, which cause degradation. In our study, the NCR is comparatively low but SNR is very good as  $SNR = \infty$ , which signifies that the host audio and the watermarked audio are absolutely the same.

Although Alhussein and Muhammad recovered the identity even at 20 dB, they did not mention that after an attack of 20dB, the speech signal of Parkinson’s disease is usable for the evaluation. In our study, we found that after the noise attack of 20dB, the audio signal cannot be diagnosed accurately. In a privacy protected healthcare system, privacy is important but not at the cost of accurate diagnosis. A comparison between the watermark algorithm of [25] and our proposed zero-watermarking algorithm is provided in Figure 13.



**Figure 13:** A comparison between the proposed algorithm and the existing algorithm.

## 5 Conclusion

A privacy protected healthcare system is developed in this study using the proposed zero-watermarking algorithm, which generates two secret shares of a subject’s identity with the help of visual cryptography. This makes the developed system more secure from unauthorized access because the identity of a subject cannot be disclosed until both shares are available simultaneously. The positive aspect of the proposed zero-watermarking algorithm is that it does not degrade the audio sample unlike traditional watermarking techniques, and hence it does not affect the accuracy of the diagnosis in the developed healthcare system. In the proposed algorithm, the secret shares of the identity are embedded into the secret keys instead of the host audio. The insertion of the secret shares depends on the patterns of the host audio, which are explored by implementing the 1D-LBP codes. The codes having the high frequency in the histogram of the speech signal are selected to embed the

secret shares of the identity. The index key holds the locations of these codes and this is used to generate the watermark key that contains the secret shares of identity. The performance of the proposed algorithm is evaluated by using the MEEI voice disorder database. The experimental results show that the proposed algorithm is reliable in the detection of a subject's identity and robust against noise attacks with various SNR. In addition, the secret shares are not embedded in the host audio, and therefore the imperceptibility of the proposed algorithm is naturally achieved.

## Acknowledgement

This work was supported by the Deanship of Scientific Research, King Saud University, Riyadh, Saudi Arabia, for funding through the Research Group under Project RG-1435-051.

## References

- [1] Z. Ali, G. Muhammad, M.F. Alhamid, An Automatic Health Monitoring System for Patients Suffering from Voice Complications in Smart Cities, *IEEE Access*, PP (2017) 1-1.
- [2] G. Muhammad, M. Alsulaiman, Z. Ali, T.A. Mesallam, M. Farahat, K.H. Malki, A. Al-nasheri, M.A. Bencherif, Voice pathology detection using interlaced derivative pattern on glottal source excitation, *Biomedical Signal Processing and Control*, 31 (2017) 156-164.
- [3] Z. Ali, I. Elamvazuthi, M. Alsulaiman, G. Muhammad, Automatic Voice Pathology Detection With Running Speech by Using Estimation of Auditory Spectrum and Cepstral Coefficients Based on the All-Pole Model, *Journal of Voice*, 30 (2016) 757.e757-757.e719.
- [4] C.-L. Hsu, M.-R. Lee, C.-H. Su, The Role of Privacy Protection in Healthcare Information Systems Adoption, *Journal of Medical Systems*, 37 (2013) 9966.
- [5] T. Gong, H. Huang, P. Li, K. Zhang, H. Jiang, A Medical Healthcare System for Privacy Protection Based on IoT, in: 2015 Seventh International Symposium on Parallel Architectures, Algorithms and Programming (PAAP), 2015, pp. 217-222.
- [6] T. Hayajneh, B. Mohd, M. Imran, G. Almashaqbeh, A. Vasilakos, Secure Authentication for Remote Patient Monitoring with Wireless Medical Sensor Networks, *Sensors*, 16 (2016) 424.
- [7] N. Roy, R.M. Merrill, S. Thibeault, R.A. Parsa, S.D. Gray, E.M. Smith, Prevalence of voice disorders in teachers and the general population, *J Speech Lang Hear Res*, 47 (2004) 281-293.
- [8] E. Ahmed, I. Yaqoob, A. Gani, M. Imran, M. Guizani, Internet-of-things-based smart environments: state of the art, taxonomy, and open research challenges, *IEEE Wireless Communications*, 23 (2016) 10-16.
- [9] Y. Mehmood, F. Ahmad, I. Yaqoob, A. Adnane, M. Imran, S. Guizani, Internet-of-Things-Based Smart Cities: Recent Advances and Challenges, *IEEE Communications Magazine*, 55 (2017) 16-24.
- [10] T. Guelzim, M.S. Obaidat, B. Sadoun, Chapter 1: Introduction and overview of key enabling technologies for smart cities and homes, in: M. S. Obaidat, P. Nicopolitidis (Eds.) *Smart Cities and Homes*, Morgan Kaufmann, Cambridge, MA 02139, USA, 2016, pp. 1-16.
- [11] L. Poissant, J. Pereira, R. Tamblyn, Y. Kawasumi, The Impact of Electronic Health Records on Time Efficiency of Physicians and Nurses: A Systematic Review, *Journal of the American Medical Informatics Association : JAMIA*, 12 (2005) 505-516.
- [12] V. McKelvey, Spending more on in-home care, Retrieved on March 1, 2017 from <http://www.aarp.org/relationships/caregiving/info-01-2010/spending-more-on-in-home-care.html> (2010).

- [13] K. Häyrynen, K. Saranto, P. Nykänen, Definition, structure, content, use and impacts of electronic health records: A review of the research literature, *International Journal of Medical Informatics*, 77 (2008) 291-304.
- [14] D. Lin, Y. Tang, F. Labeau, Y. Yao, M. Imran, A.V. Vasilakos, Internet of Vehicles for E-Health Applications: A Potential Game for Optimal Network Capacity, *IEEE Systems Journal*, 11 (2017) 1888-1896.
- [15] M.S. Hossain, Patient status monitoring for smart home healthcare, in: 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2016, pp. 1-6.
- [16] G. Muhammad, Automatic speech recognition using interlaced derivative pattern for cloud based healthcare system, *Cluster Computing*, 18 (2015) 795-802.
- [17] J. Zhou, Z. Cao, X. Dong, A.V. Vasilakos, Security and Privacy for Cloud-Based IoT: Challenges, *IEEE Communications Magazine*, 55 (2017) 26-33.
- [18] K. Zhang, J. Ni, K. Yang, X. Liang, J. Ren, X.S. Shen, Security and Privacy in Smart City Applications: Challenges and Solutions, *IEEE Communications Magazine*, 55 (2017) 122-129.
- [19] I. Yaqoob, E. Ahmed, M.H.u. Rehman, A.I.A. Ahmed, M.A. Al-garadi, M. Imran, M. Guizani, The rise of ransomware and emerging security challenges in the Internet of Things, *Computer Networks*, (2017).
- [20] S. Li, L. Da Xu, *Securing the Internet of Things*, Syngress, Elsevier Inc., Boston, United States, 2017.
- [21] F. Rahman, M.Z.A. Bhuiyan, S.I. Ahamed, A privacy preserving framework for RFID based healthcare systems, *Future Generation Computer Systems*, 72 (2017) 339-352.
- [22] S. Al-Janabi, I. Al-Shourbaji, M. Shojafar, S. Shamshirband, Survey of main challenges (security and privacy) in wireless body area networks for healthcare applications, *Egyptian Informatics Journal*, 18 (2017) 113-122.
- [23] B. Yüksel, A. Küpçü, Ö. Özkasap, Research issues for privacy and security of electronic health services, *Future Generation Computer Systems*, 68 (2017) 1-13.
- [24] C.S. Kruse, B. Frederick, T. Jacobson, D.K. Monticone, *Cybersecurity in healthcare: A systematic review of modern threats and trends*, IOS Press, 25 (2017) 1-10.
- [25] M. Alhussein, G. Muhammad, Watermarking of Parkinson Disease Speech in Cloud-Based Healthcare Framework, *International Journal of Distributed Sensor Networks*, 11 (2015) 264575.
- [26] J.W. Langston, Parkinson's Disease: Current and Future Challenges, *NeuroToxicology*, 23 (2002) 443-450.
- [27] S.V. Dhavale, R.S. Deodhar, D. Pradhan, L.M. Patnaik, A robust zero watermarking algorithm for stereo audio signals, *International Journal of Information and Computer Security*, 8 (2016) 330-346.
- [28] N. Chen, J. Zhu, A Robust Zero-Watermarking Algorithm for Audio, *EURASIP Journal on Advances in Signal Processing*, 2008 (2007) 453580.
- [29] M. Fallahpour, D. Megías, Audio Watermarking Based on Fibonacci Numbers, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23 (2015) 1273-1282.
- [30] S.R. Schwartz, S.M. Cohen, S.H. Dailey, R.M. Rosenfeld, E.S. Deutsch, M.B. Gillespie, E. Granieri, E.R. Hapner, C.E. Kimball, H.J. Krouse, J.S. McMurray, S. Medina, K. O'Brien, D.R. Ouellette, B.J. Messenger-Rapport, R.J. Stachler, S. Strode, D.M. Thompson, J.C. Stemple, J.P. Willging, T. Cowley, S. McCoy, P.G. Bernad, M.M. Patel, Clinical Practice Guideline: Hoarseness (Dysphonia), *Otolaryngology -- Head and Neck Surgery*, 141 (2009) S1-S31.
- [31] The American Heritage® Stedman's Medical Dictionary, Retrieved January 24, 2016 from Dictionary.com website <http://dictionary.reference.com/browse/dysphonia>.
- [32] T. Mau, Diagnostic evaluation and management of hoarseness, *The Medical clinics of North America*, 94 (2010) 945-960.

- [33] Q. Yan, R. Yang, J. Huang, Copy-move detection of audio recording with pitch similarity, in: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015, pp. 1782-1786.
- [34] R.H. Martins, J. Defaveri, M.A. Domingues, R. de Albuquerque e Silva, Vocal polyps: clinical, morphological, and immunohistochemical aspects, *Journal of voice : official journal of the Voice Foundation*, 25 (2011) 98-106.
- [35] J. Bohlender, Diagnostic and therapeutic pitfalls in benign vocal fold diseases, *GMS current topics in otorhinolaryngology, head and neck surgery*, 12 (2013) 1-19.
- [36] L.H. Rosenthal, M.S. Benninger, R.H. Deeb, Vocal fold immobility: a longitudinal analysis of etiology over 20 years, *Laryngoscope*, 117 (2007) 1864-1870.
- [37] K. Simonyan, F. Tovar-Moll, J. Ostuni, M. Hallett, V.F. Kalasinsky, M.R. Lewin-Smith, E.J. Rushing, A.O. Vortmeyer, C.L. Ludlow, Focal white matter changes in spasmodic dysphonia: a combined diffusion tensor imaging and neuropathological study, *Brain : a journal of neurology*, 131 (2008) 447-459.
- [38] K. Nemr, M. Simoes-Zenari, G.F. Cordeiro, D. Tsuji, A.I. Ogawa, M.T. Ubrig, M.H. Menezes, GRBAS and Cape-V scales: high reliability and consensus when applied at different times, *Journal of voice : official journal of the Voice Foundation*, 26 (2012) 812.e817-822.
- [39] S. Deguchi, Y. Ishimaru, S. Washio, Preliminary Evaluation of Stroboscopy System Using Multiple Light Sources for Observation of Pathological Vocal Fold Oscillatory Pattern, *Annals of Otology, Rhinology & Laryngology*, 116 (2007) 687-694.
- [40] R. Speyer, G.H. Wieneke, W. Kersing, P.H. Dejonckere, Accuracy of Measurements on Digital Videostroboscopic Images of the Vocal Folds, *Annals of Otology, Rhinology & Laryngology*, 114 (2005) 443-450.
- [41] M. Markaki, Y. Stylianou, Voice Pathology Detection and Discrimination Based on Modulation Spectral Features, *Audio, Speech, and Language Processing, IEEE Transactions on*, 19 (2011) 1938-1948.
- [42] Z. Ali, I. Elamvazuthi, M. Alsulaiman, G. Muhammad, Detection of Voice Pathology using Fractal Dimension in a Multiresolution Analysis of Normal and Disordered Speech Signals, *Journal of Medical Systems*, 40 (2015) 20.
- [43] Z. Ali, M. Alsulaiman, G. Muhammad, I. Elamvazuthi, A. Al-nasheri, T.A. Mesallam, M. Farahat, K.H. Malki, Intra- and Inter-database Study for Arabic, English, and German Databases: Do Conventional Speech Features Detect Voice Pathology?, *Journal of Voice*, (2017).
- [44] Z. Ali, M. Talha, M. Alsulaiman, A Practical Approach: Design and Implementation of a Healthcare Software for Screening of Dysphonic Patients, *IEEE Access* 5(2017) 5844 - 5857.
- [45] A. Al-nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, Investigation of Voice Pathology Detection and Classification on Different Frequency Regions Using Correlation Functions, *Journal of Voice*, 31 (2017) 3-15.
- [46] Massachusetts Eye & Ear Infirmary Voice & Speech LAB, *Disordered Voice Database Model 4337 (Ver. 1.03)* in, Kay Elemetrics Corp, , Boston, MA, 1994.
- [47] M. Naor, A. Shamir, Visual cryptography, in: A. De Santis (Ed.) *Advances in Cryptology — EUROCRYPT'94: Workshop on the Theory and Application of Cryptographic Techniques Perugia, Italy, May 9–12, 1994 Proceedings*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1995, pp. 1-12.
- [48] W. Abdul, Z. Ali, S. Ghouzali, B. ALfawaz, G. Muhammad, M.S. Hossain, Biometric Security Through Visual Encryption for Fog Edge Computing, *IEEE Access*, 5 (2017) 5531 – 5538.
- [49] R.L. Devaney, *The Gingerbreadman*, *Algorithm*, 3 (1992) 15-16.
- [50] L. Houam, A. Hafiane, A. Boukrouche, E. Lespessailles, R. Jennane, One dimensional local binary pattern for bone texture characterization, *Pattern Analysis and Applications*, 17 (2014) 179-193.
- [51] N. Chatlani, J.J. Soraghan, Local binary patterns for 1-D signal processing, in: 2010 18th European Signal Processing Conference, 2010, pp. 95-99.



- [52] J.D. Arias-Londoño, J.I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, G. Castellanos-Domínguez, An improved method for voice pathology detection by means of a HMM-based feature space transformation, *Pattern Recognition*, 43 (2010) 3100-3112.
- [53] A. Gelzinis, A. Verikas, M. Bacauskiene, Automated speech analysis applied to laryngeal disease categorization, *Computer methods and programs in biomedicine*, 91 (2008) 36-47.
- [54] T.A. Mesallam, M. Farahat, K.H. Malki, M. Alsulaiman, Z. Ali, A. Al-nasheri, G. Muhammad, Development of the Arabic Voice Pathology Database and Its Evaluation by Using Speech Features and Machine Learning Algorithms, *Journal of Healthcare Engineering*, 2017 (2017) 13.
- [55] A. Zulfiqar, A. Muhammad, A.M. Martinez-Enriquez, G. Escalada-Imaz, Text-independent speaker identification using VQ-HMM model based multiple classifier system, in: *Advances in Soft Computing*, Springer Berlin Heidelberg, 2010, pp. 116-125.
- [56] Y. Lin, W.H. Abdulla, Principles of Psychoacoustics, in: *Audio Watermark: A Comprehensive Foundation Using MATLAB*, Springer International Publishing, Cham, 2015, pp. 15-49.
- [57] B.E. Boser, I.M. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, in: *Computational learning theory*, 5th annual workshop on, ACM, 1992, pp. 144-152.
- [58] C. Cortes, V. Vapnik, Support-Vector Networks, *Machine Learning*, 20 (1995) 273-297.

## Appendix A

List of patients recorded at sampling frequency of 50 KHz in MEEI voice disorder database

No.	Patient's Identity	Diagnosis	No.	Patient's Identity	Diagnosis
1	AJF12AN	<sup>M</sup> A-P squeezing	27	JAJ10AN	N/A
2	AJM05AN	Hyper function	28	JCH21AN	<sup>M</sup> Paralysis
3	AMT11AN	N/A	29	JEC18AN	N/A
4	BJK16AN	N/A	30	JES29AN	N/A
5	CAK25AN	<sup>M</sup> Vocal fold edema	31	JJD06AN	N/A
6	CRM12AN	<sup>M</sup> Parkinson's Disease	32	JJD11AN	Vocal fold edema
7	CXM18AN	N/A	33	JLS11AN	<sup>M</sup> Vocal fold polyp
8	DAC26AN	<sup>M</sup> Paralysis	34	JMH22AN	<sup>M</sup> Vocal fold polyp
9	DBF18AN	<sup>M</sup> Vocal fold edema	35	JPP27AN	<sup>M</sup> Paralysis
10	DLW04AN	N/A	36	JTM05AN	Reinke's edema
11	DMP04AN	<sup>M</sup> A-P squeezing	37	JWK27AN	N/A
12	DWK04AN	<sup>M</sup> Hyper function	38	KAH02AN	N/A
13	DXS20AN	N/A	39	KAS09AN	<sup>M</sup> Hyper function
14	EAB27AN	<sup>M</sup> Cyst	40	KLC06AN	<sup>M</sup> Paralysis
15	EAL06AN	N/A	41	KMS29AN	<sup>M</sup> Paralysis
16	EJM04AN	<sup>M</sup> Hyper function	42	LAR05AN	<sup>M</sup> A-P squeezing
17	EML18AN	<sup>M</sup> Idiopathic dysphonia	43	LGM01AN	<sup>M</sup> Vocal fold edema
18	EWV05AN	<sup>M</sup> Hyper function	44	LJH06AN	<sup>M</sup> A-P squeezing
19	EXE06AN	<sup>M</sup> Vocal fold edema	45	LMB18AN	N/A
20	EXI04AN	<sup>M</sup> Hyper function	46	LPN14AN	<sup>M</sup> A-P squeezing
21	EXI05AN	<sup>M</sup> A-P squeezing	47	LSB18AN	N/A
22	EXW12AN	N/A	48	LXG17AN	N/A
23	FGR15AN	<sup>M</sup> Hyper function	49	LXY01AN	<sup>M</sup> Paralysis
24	FMM29AN	N/A	50	MAT26AN	Hyper function
25	HWR04AN	<sup>M</sup> A-P squeezing	51	MAT28AN	N/A
26	HXI29AN	<sup>M</sup> A-P squeezing	52	MCB20AN	<sup>M</sup> A-P squeezing

(Continue) List of patients recorded at sampling frequency of 50 KHz in MEEI voice disorder database

No.	Patient's Identity	Diagnosis	No.	Patient's Identity	Diagnosis
53	MMS29AN	<sup>M</sup> A-P squeezing	66	RXM15AN	<sup>M</sup> A-P squeezing
54	MRC20AN	A-P squeezing	67	RXS13AN	<sup>M</sup> Keratosis
55	MRM16AN	N/A	68	SAR14AN	<sup>M</sup> A-P squeezing
56	MWD28AN	<sup>M</sup> Hyper function	69	SEC02AN	<sup>M</sup> Vocal nodules
57	MYW14AN	N/A	70	SHC07AN	<sup>M</sup> Vocal nodules
58	NJS06AN	<sup>M</sup> Vocal nodules	71	SLG05AN	Hyper function
59	NMB28AN	<sup>M</sup> Keratosis	72	TAB21AN	<sup>M</sup> Hyper function
60	PAT10AN	<sup>M</sup> A-P squeezing	73	TLP13AN	<sup>M</sup> Vocal nodules
61	PDO11AN	<sup>M</sup> Cyst	74	TMD12AN	<sup>M</sup> Vocal tremor
62	PMC26AN	Varix	75	TPP11AN	N/A
63	REC19AN	<sup>M</sup> Hyper function	76	WJP20AN	<sup>M</sup> Hyper function
64	RMF14AN	<sup>M</sup> Hyper function	77	WXE04AN	<sup>M</sup> A-P squeezing
65	RPJ15AN	<sup>M</sup> A-P squeezing			

<sup>1</sup>**Note:** N/A represents the patients whose diagnoses are not provided in the documentation of the MEEI database.

<sup>2</sup>**Note:** M in the superscript denotes that the patient is suffering from more than one disorder. However, only one disorder is listed in the above table.