Received March 20, 2017, accepted April 6, 2017, date of publication April 13, 2017, date of current version May 17, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2693282

A Practical Approach: Design and Implementation of a Healthcare Software for Screening of Dysphonic Patients

ZULFIQAR ALI[1], MUHAMMAD TALHA[2], AND MANSOUR ALSULAIMAN[1]
[1]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[2]Deanship of Scientific Research, King Saud University, Riyadh 11543, Saudi Arabia

Corresponding authors: Zulfiqar Ali (zuali@ksu.edu.sa) and Muhammad Talha (mnaseem@ksu.edu.sa)

This work was supported by the Deanship of Scientific Research, King Saud University, Riyadh, Saudi Arabia, through the Research Group under Project RG-1437-037.

ABSTRACT Risk management in the development of medical software and devices is one of the most crucial processes in ensuring accurate diagnoses and treatment of disease. The consequences of wrong decisions that happen in our daily life might be unembellished. However, wrong decisions in healthcare based on unreliable evidence due to erroneous software could result in loss of life. Dysphonic patients suffering from various vocal fold disorders might have a threat of life due to inaccurate diagnosis. Some voice disorders, such as keratosis, are precancerous, and can become cancerous in cases that involve inaccurate diagnosis due to software failure. The objective of this paper is to design and implement a healthcare software for the detection of voice disorders in nonperiodic speech signals. Occurrences of potential risks during the design and development of the proposed software are taken into account to avoid failure. The software is implemented by applying the local binary pattern (LBP) operator on the textures of nonperiodic signals. The textures are obtained through the recurrence plot. The LBP operator computes the histograms for normal persons and dysphonic patients, and these histograms are used with the support vector machine for the automatic classification of dysphonic patients. The software is evaluated and tested by using the Massachusetts Eye and Ear Infirmary voice disorder database. The success rate of the proposed healthcare system is 97.73% $\pm$ 1.2, and the area under the receiver operating characteristic curve is 0.98 $\pm$ 0. The performance of the proposed healthcare system is much better than the existing commercial software used for screening dysphonic patients.

INDEX TERMS Risk management, vocal fold disorders, recurrence plot, local binary pattern, type 2 and 3 signals.

I. INTRODUCTION

Medical devices operated with software are very important in healthcare [1]–[3], and they provide important complimentary information to the clinicians in making accurate decisions and treatment plans. In the future, however, the recall of medical devices due to software and hardware malfunctions may rise to an alarming level [4]–[6]. Medical care is a critical area of safety, and systems failures in this area can cause threats to life or damage to the environment. It is important that the quality of software should be high in safety critical devices by following the standards. According to Software Product Liability report [7], incorrect matching of patient and data, faulty programming, erroneous calculation, and many other software errors have resulted in the recall of medical devices. Rakitin [8] mentions that it is a responsibility of the companies to show that their developed software is safe.

Faulty software increases the risks faced by patients, which should be reduced through the risk management process. Richard [9] defines the risk management as a step-by-step process to identify and handle risk factors. Boehm [10] states that risk management is not a cookbook approach, but rather it includes the initial identification of risks, handling of those risks, and continuous management of them. The objective of this study is to design and develop a healthcare software that will screen dysphonic patients. The software is based on the local binary pattern (LBP) features, which are independent of fundamental frequency (F0). The features depending on the F0 are one of the major risks in the failure of disorder detection software, especially when the input signal is non-periodic.

The voice of a person can be considered to be healthy if he/she can meet the personal and professional requirements of the voice in daily life without facing any fatigue and vocal problems [11]. Whilst anybody can be affected by a voice disorder, certain professionals such as teachers, singers, and lawyers experience more risks of suffering from a voice disorder [12], [13]. People suffering from various types of voice disorders are referred to as dysphonic patients. Generally, vocal misuse (e.g., yelling, excessive talking, screaming, and crying) can cause irritating forces at the contact place of two vocal folds. Other than these reasons, some other factors such as poor hydration, medication, alcohol consumption, and smoking also contribute towards and at times directly influence the development of vocal fold disorders [6], [7]. Therefore, damaged vocal folds exhibit abnormal vibrations during voice production. Any change along the vocal folds may disturb the vibrations of the vocal fold, and affect the quality of the voice. According to Titze [14], "*The qualitative change in the behavior of a dynamical system is known as a bifurcation. It usually occurs when some parameter of the vibrating system is changed gradually (e.g., lung pressure, vocal fold tension, or asymmetry between the vocal folds).*" A speech signal can be classified into three types based on the bifurcation: 1) Type 1 are nearly periodic, 2) Type 2 exhibit qualitative changes (bifurcation) and, therefore, the F0 of the signal changes over time, and 3) Type 3 are non-periodic with significant noise components. Based on the recommendation of Titze, the acoustics measures depending on the perturbation analysis are not reliable for Type 2 and Type 3 signals. The most commonly used perturbation measures are shimmer and jitter [15], [16]. Jitter represents cycle-to-cycle frequency perturbation and shimmer refers to cycle-to-cycle amplitude perturbation. Both measures are strongly dependent on the accurate estimation of F0, and it is itself a challenging task in the case of Type 2 and Type 3 signals [17]. Therefore, in some studies [18], [19] that use F0 dependent acoustic measures, the signals of Type 3 are not included for the detection of voice disorders. Type 3 of the signals degraded the accuracy of the system developed in [18] by 10%. So, features not relying on F0 can be a better choice for the classification of dysphonic patients by using Type 2 and Type 3 signals.

F0 independent features can be extracted from the recurrence plots [20]–[22] of speech signals. Recurrence plots provide visual facts of a time series and a two-dimensional image is obtained as a result. Recurrence plots compute the resemblance of a point in time-series data to all other points [23]. The recurrence plot of a signal that is obtained by the addition of two sine waves is depicted in Fig. 1(b), in which the frequency of the first wave is 200 Hz and the second is 400 Hz. The sampling frequency is 8 KHz for both sine waves, and the signal itself is depicted in Fig. 1(a) and referred to as a clean signal. Moreover, a white noise of signal-to-noise ratio (SNR) of 20 dB is added to the clean signal, and the resultant signal is depicted in Fig. 1(c), referred to as a noisy signal. The recurrence plot of the noisy signal is shown in Fig. 1(d), and it can be observed that the recurrence plot of the noisy signal is blurred when compared with that of the clean signal. The difference between the two plots can be noticed in Fig. 1(b) and Fig. 1(d). This fact can be used for the classification of normal and disordered subjects. The speech signal of a normal subject is considered as periodic and does not contain any noise generated during the production of voice. On the other hand, the speech signal of a disorder subject contains noisy components due to irregular vibrations of the disordered vocal folds. Voice disorders make the voice strained, hard, weak, whispering, and breathy [24], and they generate the noisy component in voice disordered patients. Therefore, the speech produced by dysphonic patients sounds unpleasant.

The motivation behind the development of the proposed software is automatic screening of dysphonic patients without any potential risk. Automatic evaluation of voice disorders has the advantage over the subjective evaluation due to the following limitations: size of the assessment panel [25], human error, attention, memory lapses of raters [25], [26], professional background of raters [27], and disagreement of judgement between slight and moderate voice disorders [25], [26], [28]. Such software can be used for the early screening of a patient to avoid complications, as some voice disorders such as keratosis become life threatening [29]. Furthermore, healthcare software for the screening of dysphonic patients can be deployed in remote areas where a general practitioner can evaluate the patient and refer him/her to a specialized clinic.

The healthcare software that is proposed in this study classifies dysphonic patients by using non-periodic speech signals of Type 2 and Type 3. This is a challenging task, but the existing software fails to provide high accuracy for these types of signals due to the usage of F0 dependent features [18], [19]. To avoid the F0 dependent features, new features are used to develop the proposed software. The new features in the proposed software are computed by using textures of recurrence plots and LBP operators [30]. These features have never been used before in the screening of dysphonic patients. To enhance the efficiency of the proposed software, the number of computed features is reduced by applying uniform mapping [31] on LBP codes. Uniform mapping reduces the number of features as well as improves the accuracy of the system [30], [32], [33]. Moreover, to determine the most discriminant features when screening for dysphonic patients, Fisher's Discrimination Ratio (FDR) is used. In addition, the developed software uses support vector machine (SVM) for automatic identification of dysphonic patients. Several experiments are conducted for the evaluation of the proposed software. To avoid the bias of training and testing data sets, k-folds cross validation is implemented during evaluation of the software. The accuracy of the proposed software in the detection of voice disordered patients by using non-periodic signal as an input is very good and better than the existing software [34].

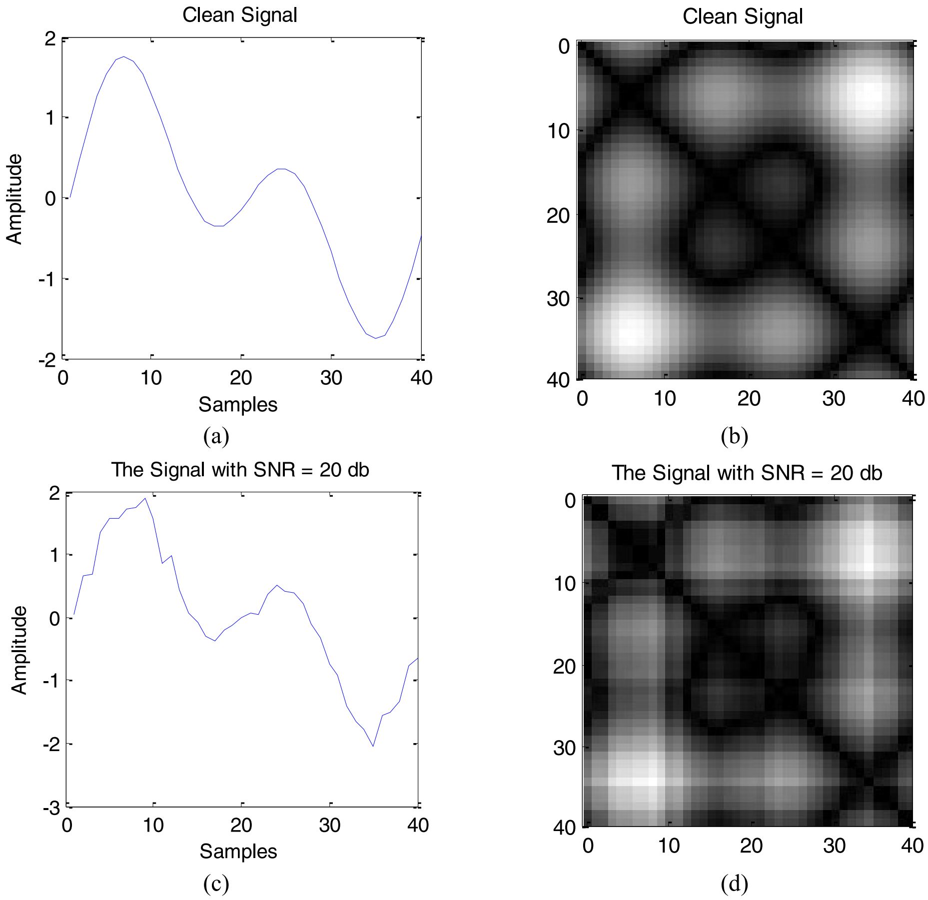

FIGURE 1. Signals and their recurrence plots: (a) Clean signal, (b) Recurrence plot for clean signal, (c) Noisy signal (clean signal + white noise of SNR = 20 dB), and (d) Recurrence plot for noisy signal.

The rest of the paper is organized as follows: Section II describes the design and implementation of the proposed healthcare software for the screening of dysphonic patients with potential risks. Section III conducts an extensive evaluation of the proposed software by using the clinical data. The experiments are performed by using k-folds cross validation with the full set of the computed features, as well as a subset containing the most discriminant features. Section IV provides a discussion on the developed software and its deployment. Finally, Section V concludes the study.

II. DESIGN AND IMPLEMENTATION OF THE PROPOSED SCREENING SOFTWARE

A. DESIGN OF THE PROPOSED SOFTWARE

The design of the proposed software is presented in Fig. 2. Various potential risks are considered during the design of the proposed software, such as the features are calculated in a way that they do not rely on F0. The reason is that the developed software works with non-periodic signals of Type 2 and Type 3, and accurate estimation of F0 is not possible for these types of signals. The first step in the computation of these new types of features is a generation of the textures of normal subjects and disordered subjects of Type 2 and 3 by using a recurrence plot. The texture of both subjects is analyzed by applying the LBP operator, and histograms are obtained as a result. The histograms contained 256 bins, and a uniform mapping is applied to reduce the number of bins to 59 for the time efficiency of the proposed healthcare software. The first of the three computed histograms is Sign-LBP, and the second is Mag-LBP. Both types of the histograms are described in subsection *Recurrence Plot and Uniform LBP*. The third histogram is a concatenation of Sign- and Mag-LBP, and referred to as ConcSM-LBP. All these histograms, Sign-LBP, Mag-LBP, and ConcSM-LBP, are computed for each normal and disordered subject. The number of features (bins) in Sign- and Mag-LBP is 59, while there are 118 features in ConcSM-LBP. To reduce the number of features, FDR is used and the top three features with the greater ratio are selected.

The automatic detection of dysphonic patients is achieved by providing the computed histograms to the SVM one by one. Another type of risk that can occur during the evaluation of the proposed software is a biasness of the software due to the training and testing dataset. To overcome this risk, the k-fold cross validation approach is used during the implementation of the software.

B. IMPLEMENTATION OF THE PROPOSED SOFTWARE

During implementation of the proposed software, the first component is a selection of non-periodic signals of Type 2 and Type 3 from the voice disorder database. The second

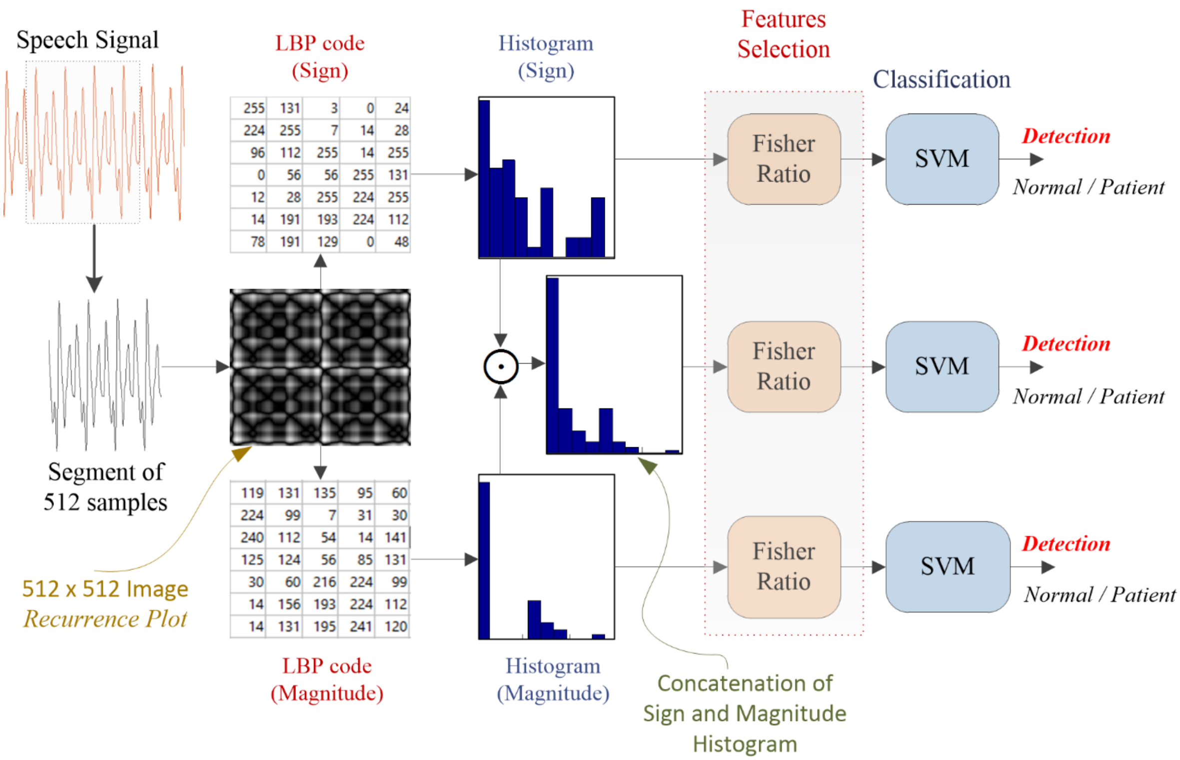

FIGURE 2. The block diagram of the proposed screening software.

component is the extraction of new types of features by using recurrence plots and LBP operator. The third component is FDR, which determines the most discriminant features for screening dysphonic patients. The fourth component is the use of SVM for the automatic detection of dysphonic patients. These components are described in the following subsections.

1) SELECTION OF TYPE 2 AND TYPE 3 SIGNALS

Type 2 and Type 3 signals are taken from the voice disorder database [35] recorded in the lab at the Massachusetts Eye and Ear Infirmary (MEEI), and openly commercialized by Kay Elemetrics. The database contained recorded speech samples of the sustained vowel /ah/ vocalized by dysphonic patients as well as normal persons. The speech samples of Type 2 and Type 3 signals are selected from the subset of the MEEI database according to the criteria described by Titze in [14]. The subset contained 173 disordered and 53 normal samples and has been used in a number of studies [36]–[41]. Type 3 signals are non-periodic in nature and contain noisy components, whereas Type 2 signals have strong variation in F0 due to bifurcation. Some Type 2 and Type 3 speech samples are plotted in Fig. 3, and the selected samples will be used for the evaluation of the proposed healthcare software.

An analysis is performed for Type 2 and Type 3 signals by computing F0, standard deviation of F0 (stdF0), shimmer and jitter. Box plots for stdF0, shimmer, and jitter are plotted in Fig. 4. It can be observed that the range of the stdF0 is 157.3 and the mean stdF0 is 12.5 with standard deviation (STD) of 27.0. A large value of mean and STD shows that these types of signals vary quickly over time, and especially in signals of Type 3, F0 is hard to estimate accurately due to noisy components. Moreover, the amplitude perturbation of both types of signal is analyzed by observing the shimmer. The statistics of shimmer are: range 25, median 7.5, mean 9.1, and STD 5.6. The mean and STD of shimmer are large, and it shows that these signals observe a lot of perturbations in amplitude. The statistics of jitter also suggest that Type 2 and Type 3 signals have strong frequency perturbations. The mean and STD of jitter are 3.5 and 3.9, respectively.

2) RECURRENCE PLOT AND UNIFORM LBP

The recurrence plot, RP, for a speech signal $S = s_1, s_2, s_3, \ldots, s_n$ is computed as

$$RP_i = abs\,(S - s_i) \text{ where } i = 1, 2, 3, \ldots, n \tag{1}$$

The obtained RP is an image of $n \times n$ dimension, and i stands for the row index of RP.

After getting the RP, each element of it is replaced by an LBP code. To determine the LBP code, a 3×3 window is centered on each element of the RP, and the center element of the selected window is represented by r_c and the remaining eight neighbors by $r_1, r_2, r_3, \ldots, r_8$, where r_1 is the element at the right bottom corner of the r_c and other elements $r_1, r_2, r_3, \ldots, r_8$ are taken clockwise starting from r_1. The elements $r_1, r_2, r_3, \ldots, r_8$ in Fig. 5 are 8, 4, 4, 7, 4, 2, 7, and 6, respectively, and the center element r_c is 5.

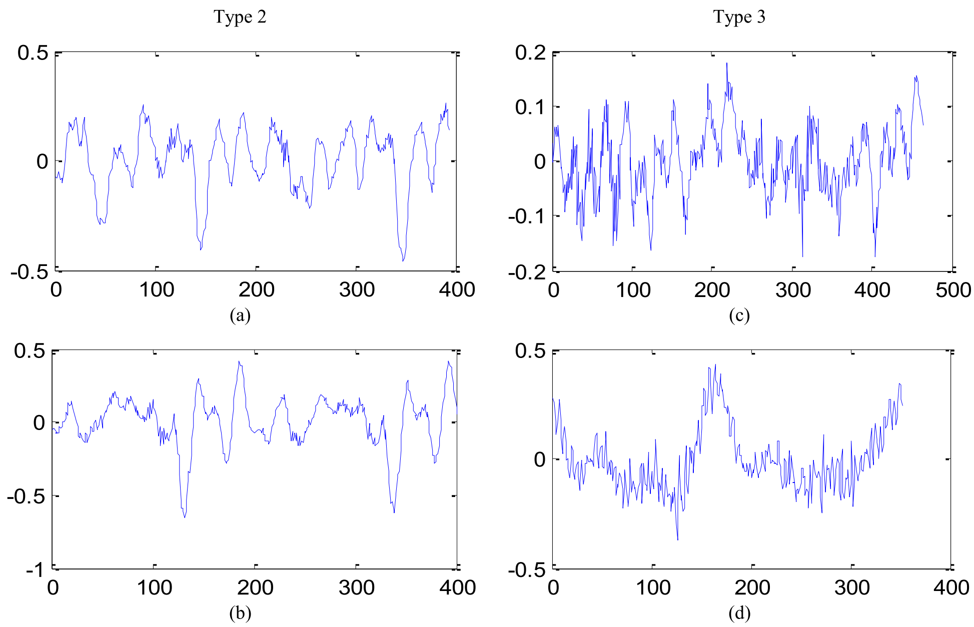

FIGURE 3. Some examples of Type 2 and Type 3 speech signals from MEEI database: (a) SAV18AN.NSP, (b) WJP20AN.NSP, (c) PAT10AN.NSP, and (d) SCC15AN.NSP.

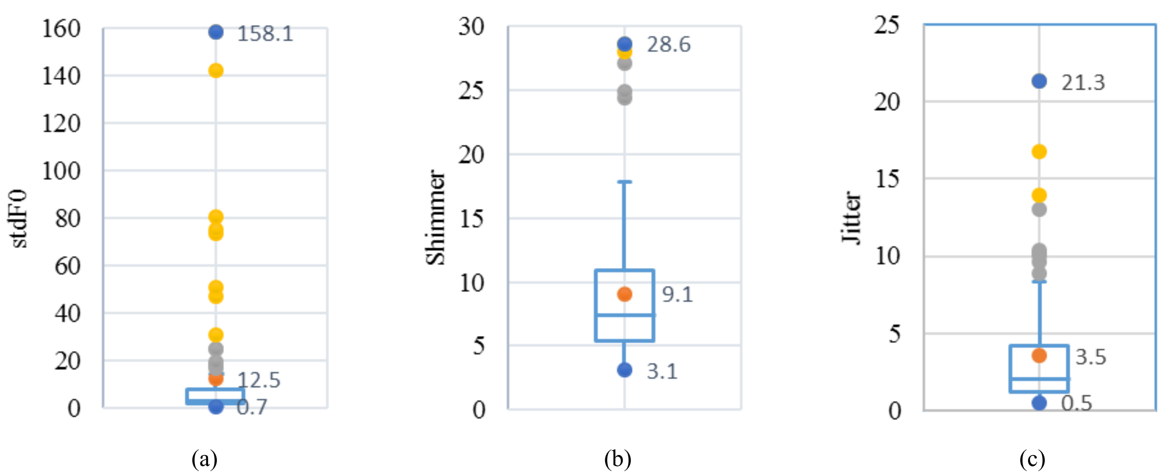

FIGURE 4. Box plots for (a) Standard deviation of F0 (stdF0), (b) Shimmer, and (c) Jitter.

The neighboring elements are now compared with the center element, and if a neighbor is equal to or larger than the center element then the neighbor will be replaced by 1, otherwise, it will be replaced by 0. Similarly, each neighboring element will be compared and replaced with 1 or 0. The calculations are done by using Eq. 2, and a binary number $b_7b_6b_5b_4b_3b_2b_1b_0$ is obtained, where b_7 is a most significant bit (MSB) and b_0 is a least significant bit (LSB).

$$b_{k-1} = \begin{cases} 1 & if\ r_k \geq r_C \\ 0 & if\ r_k < r_C \end{cases} \quad where\ k = 1, 2, 3, \ldots, 8 \tag{2}$$

Then, the binary number $b_7b_6b_5b_4b_3b_2b_1b_0$ is converted to a decimal number by multiplying each bit with a power of two by using Eq. 3.

The obtained decimal number is a required LBP code, and it will replace the center element. In a similar way, an LBP code is computed for each element of RP. This LBP code is referred to as Sign-LBP because it depends on the sign of the difference of the center and neighboring elements.

$$LBP = \sum_{k=1}^{8} \left(b_{k-1} \times 2^{k-1} \right) \tag{3}$$

To calculate the magnitude LBP, the whole procedure is the same, with the exception that the center element of the 3×3 window will be computed as

$$M = \frac{1}{8} \sum_{k=1}^{8} |r_k| \tag{4}$$

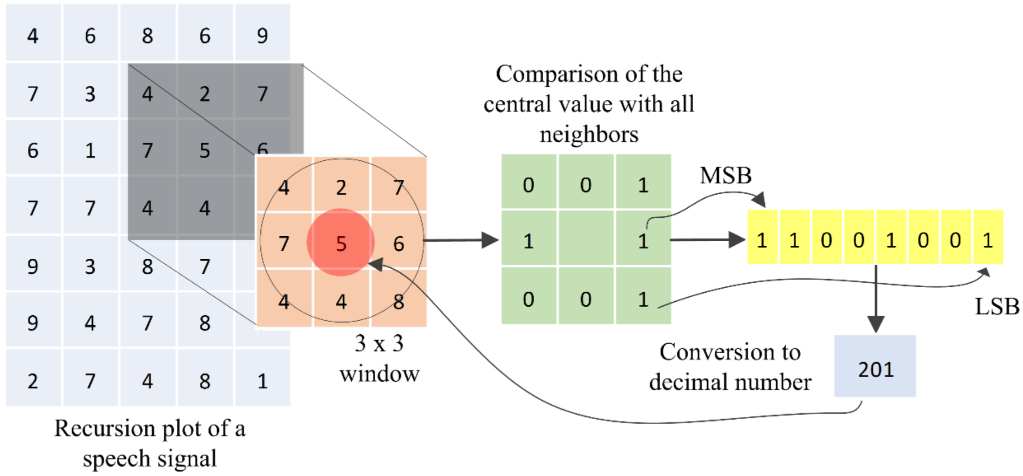

FIGURE 5. Steps to compute the LBP code.

where M represents the average magnitude of the 3×3 window. Each neighbor is compared with the M to determine the magnitude LBP code, which is referred to as Mag-LBP.

The range of both Sign- and Mag-LBP is 0 to 255, and therefore the histogram of Sign-LBP and Mag-LBP contained 256 bins, where each one describes the frequency of an LBP code. To reduce the number of bins, uniform mapping is applied on LBP codes. An LBP code is considered uniform if it has a maximum two transitions of 0-to-1 or 1-to-0. For instance, 00111100, 11110000, 00000000, and 11100111 are uniform codes, whereas 00110101, 01100110, and 10101010 are non-uniform codes as they have five, four, and seven transitions, respectively. All non-uniform codes are assigned to a single histogram bin, while each uniform code is assigned to a separate histogram bin. The number of uniform codes for eight neighbors are 58. After uniform mapping, the histogram will have 59 bins, 58 for uniform codes, and one for non-uniform codes.

3) FISHER's DISCRIMINANT RATIO

The Fisher ratio is applied to determine the features that can contribute significantly to the detection of dysphonic patients, and is given by

$$FDR_i = \frac{(\mu_{Ni} - \mu_{Di})^2}{\sigma_{Ni}^2 + \sigma_{Di}^2} \; where\; i = 1, 2, 3, \ldots, l \qquad (5)$$

where l represents the number of bins, μ_{Ni} and μ_{Di} represent the mean of the i^{th} bin of all histograms of normal and disordered subjects, respectively, while the variance of the i^{th} bin of all histograms of normal and disordered subjects is given by σ_{Ni} and σ_{Di}.

The ratio will be greater if the difference between the means of a bin of all normal subjects and all corresponding bins of disordered subjects is large, and at the same time the bins of both subjects should have small variance. The bin with the greater Fisher ratio will be more discriminant than others.

4) SUPPORT VECTOR MACHINE

SVM is a supervised learning algorithm as it needs training data to learn [42], [43]. On the basis of learning from training data, SVM can predict the class of an object, and is considered to be one of the most successful classification approaches. SVM has been applied in many real-life problems to maximize the distance between the classes. In the proposed healthcare software, SVM is used to classify the normal and pathological subjects, where pathological and normal subjects are considered to be positive and negative classes, respectively. The ultimate goal of the SVM is to find an optimal hyperplane that provides a maximum distance between the instances of the two classes.

In the case when data are not linearly separable, kernel function is implemented to map the original input space to higher dimensional space, where features are linearly separable. In this study, classification of pathological and normal samples is carried out by using LIBSVM [44] with the radial basis function (RBF), which is given by Eq. 6.

$$K\left(x, x'\right) = \exp\left(-\gamma \left\|x - x'\right\|^2\right) \qquad (6)$$

where x is the training sample, x' is the testing sample, and γ is a free parameter.

III. TESTING AND EVALUATION OF THE PROPOSED SOFTWARE

The MEEI database is used for testing and evaluating the developed healthcare software, and is recorded at two different sampling frequencies: 25 KHz and 50 KHz. Therefore, all selected Type 2 and 3 speech signals are down sampled at 25 KHz to have a unique sampling frequency. The duration of recorded samples of the sustained vowel /ah/ for normal subjects is 3 seconds, and for disordered subjects it is 1 second. Due to the difference in the duration of normal and pathological samples, only two frames of length ≈20 milliseconds (512 samples) are considered from the

TABLE 1. Screening results of the proposed healthcare software by using the full set of features.

LBP	Kernel	Sensitivity ± STD	Specificity ± STD	Accuracy ± STD	AUC ± STD
Sign-LBP	*Linear*	98.85 ± 2.1	92.52 ± 6.9	**96.21 ± 1.3**	0.98 ± 0
	Quadratic	100 ± 0	88.99 ± 9.8	95.45 ± 3.9	0.95 ± 0
	Cubic	100 ± 0	86.98 ± 7.5	94.7 ± 3.5	0.95 ± 0
	RBF	98.72 ± 2.2	92.37 ± 6.6	96.21 ± 3.5	0.97 ± 0
Mag-LBP	*Linear*	97.53 ± 4.3	92.27 ± 3.5	**95.45 ± 2.3**	0.96 ± 0
	Quadratic	98.81 ± 2.1	86.4 ± 6.5	93.94 ± 1.3	0.89 ± 0
	Cubic	100 ± 0	86.02 ± 7.7	94.7 ± 2.6	0.95 ± 0
	RBF	97.7 ± 4	88.79 ± 10.2	93.94 ± 1.3	0.94 ± 0.1
ConcSM-LBP	*Linear*	97.36 ± 2.3	90.3 ± 3.5	94.7 ± 1.3	0.95 ± 0
	Quadratic	98.72 ± 2.2	88.52 ± 5	94.7 ± 2.6	0.95 ± 0
	Cubic	100 ± 0	89.65 ± 4.8	**96.21 ± 1.3**	0.95 ± 0
	RBF	97.44 ± 4.4	89.25 ± 9.2	94.7 ± 2.6	0.96 ± 0

middle of the sustained vowel, and then recurrence plots of dimension 512×512 are generated. The other reason to select the sample of 20 milliseconds is to avoid the computational cost in the calculation of LBP codes.

All experiments for evaluation of the proposed software are performed by using the three-fold cross validation approach. Each time one of the distinct folds is used for the testing, the remaining two folds are used for the training. The results of the software evaluation are provided in the form of the following performance measures: *sensitivity*, which is the likelihood of the system to detect a disordered subject when an input signal is also a disordered subject; *specificity*, which is the likelihood of a system to detect a normal subject when an input signal is also a normal subject; *accuracy*, which is the ratio of truly classified normal and disordered subjects with the total number of subjects; and area under the Receiver Operating Characteristic (ROC) curve. A ROC curve graphically represents the quality of a classifier in the differentiation of two classes. The area under the ROC curve (AUC) closest to one represents that a classifier is capable of differentiating the two classes significantly, and it is reliable in the decision. The sensitivity (SEN), specificity (SPE), and accuracy (ACC) measures are defined by Eqs. 7, 8, and 9, respectively.

$$SEN = \frac{TP}{TP + FN} \times 100 \tag{7}$$

$$SPE = \frac{TN}{TN + FP} \times 100 \tag{8}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{9}$$

where TP, TN, FP, and FN stand for True Positive, True Negative, False Positive, and False Negative, respectively. TP means a disordered subject is classified as disordered, TN means a normal subject is detected as normal, FN means a disordered subject is misclassified as normal, and FP means a normal subject is misclassified as disordered. The SVM is implemented with linear, quadratic, cubic, and RBF kernels.

A. EVALUATION OF THE PROPOSED SOFTWARE BY USING THE FULL SET OF FEATURES

The full set of features contains 59 bins of Sign-LBP, 59 bins of Mag-LBP, and 118 bins of ConcSM-LBP. The Sign-LBP obtained a maximum ACC of 96.21%, where the corresponding SEN is 98.85%, which suggests that Sign-LBP is good at classifying normal subjects. The SPE of 92.52% shows that the classification of a disordered subject is also good with Sign-LBP. The AUC for Sign-LBP is 0.98, which indicates that the obtained results of the classifier are reliable. The ACC of 96.21% for Sign-LBP with a linear kernel is better than with quadratic, cubic, or RBF kernels, which means that normal and disordered subjects are linearly separable. The lower STD of 1.3 shows that the margin of error in the ACC of the proposed healthcare software is small.

The classification ACC for Mag-LBP is 95.23%, whereas SEN and SPE are 97.53% and 92.27%, respectively. The best ACC for Mag-LBP is also achieved with a linear kernel, which shows that two classes are also linearly separable in the case of Mag-LBP. The STD is comparatively larger than Sign-LBP, and hence the margin of error for ACC is a bit large. The AUC is 0.96, which is also good.

When two histograms of Sign-LBP and Mag-LBP are concatenated into ConcSM-LBP, the best obtained ACC is 96.21% with a STD of 1.3. In that case, the total number of features becomes 118 (=59 + 59), as each histogram has 59 bins. Normal and disordered subjects are not linearly separable for ConcSM-LBP, and this is why the best ACC for ConcSM-LBP is obtained with a cubic kernel. It can be observed from Table 1 that the corresponding SEN and SPE are 100% and 89.65%, respectively. The SPE is not good for ConcSM-LBP, and its STD is also very high at 4.8.

A comparison of performance measures is depicted in Fig. 6, in which the SEN, SPE, ACC, and AUC are plotted with error bars, which highlight the lower and upper limits of the performance measures. The error bars are plotted using the STD of the measures. Overall, maximum ACC is obtained

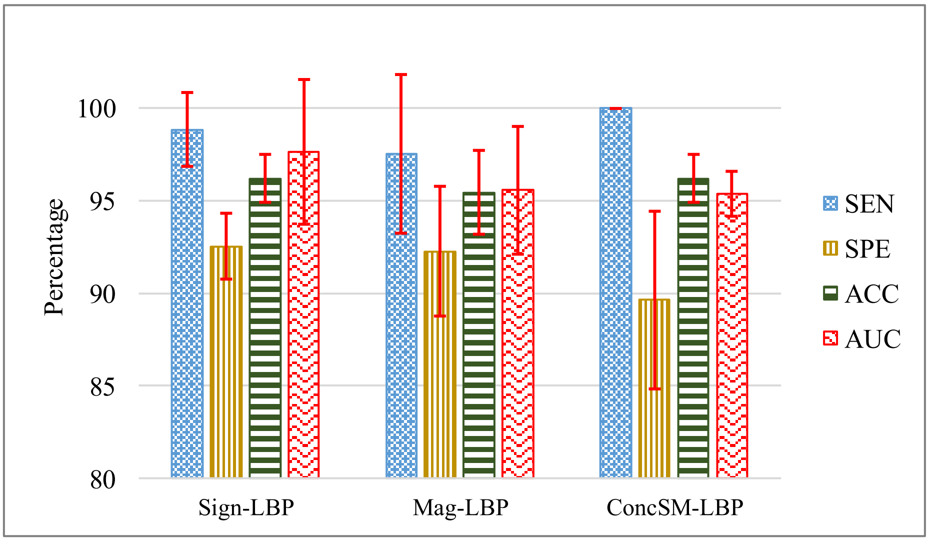

FIGURE 6. A comparison of performance measures for the full set of features.

TABLE 2. Screening results of the proposed healthcare software by using feature selection.

LBP	Kernel	Sensitivity ± STD	Specificity ± STD	Accuracy ± STD	AUC ± STD
Sign-LBP	*Linear*	95.15 ± 5.4	94.44 ± 5.6	94.7 ± 1.3	0.94 ± 0
	Quadratic	96.1 ± 4	76.25 ± 2.8	88.64 ± 2.3	0.87 ± 0
	Cubic	97.38 ± 2.3	90.41 ± 2.1	94.7 ± 2.6	0.96 ± 0
	RBF	95.95 ± 4.4	97.78 ± 3.8	**96.97 ± 2.6**	0.97 ± 0
Mag-LBP	*Linear*	96.23 ± 0.3	90.42 ± 3.1	93.94 ± 1.3	0.95 ± 0.1
	Quadratic	100 ± 0	0 ± 0	61.36 ± 4.5	0.5 ± 0.1
	Cubic	97.33 ± 2.3	87.85 ± 10.9	93.94 ± 4.7	0.93 ± 0.1
	RBF	96.47 ± 3.5	91.08 ± 7.9	**93.94 ± 1.3**	0.97 ± 0
ConcSM-LBP	*Linear*	98.72 ± 2.2	92.27 ± 3.5	96.21 ± 1.3	0.98 ± 0
	Quadratic	95.06 ± 4.3	92.48 ± 8.5	93.94 ± 1.3	0.94 ± 0
	Cubic	98.81 ± 2.1	94.2 ± 0.5	96.97 ± 1.3	0.97 ± 0
	RBF	98.61 ± 2.4	96.48 ± 3.1	**97.73 ± 1.2**	0.98 ± 0

by Sign-LBP and ConcSM-LBP, which is 96.21% with an STD of 1.3. The results obtained with Sign-LBP are more reliable as it has greater AUC than ConcSM-LBP. Moreover, in the case of Sign-LBP, the features are lineally separable.

B. EVALUATION OF THE PROPOSED SOFTWARE BY USING THE FEATURE SELECTION

In the previous section, the software was evaluated using all features (59 bins of histograms). The number of features becomes 118 when histograms of Sign- and Mag-LBP are concatenated. In other words, a normal and pathological sample is represented by 118 features, which is a large number. Therefore, feature selection is achieved by applying FDR to determine the most significant bins of the histograms for the classification of normal and disordered subjects. In this section, the evaluation of the proposed software is performed by selecting the top three features of Sign-LBP, Mag-LBP, and ConcSM-LBP. The top three features are selected by sorting the FDR of the bins in descending order. The results of the evaluation using feature selection are provided in Table 2.

The maximum obtained ACC for Sign-LBP with the top three features is 96.97%. The SEN and SPE are 95.95% and 97.78%, respectively, which is really good. Sign-LBP classified the normal as well as disordered subjects with high ACC. Moreover, a large value of AUC (i.e. 0.97) shows the reliability of the proposed software. The result is obtained with a RBF kernel, which maps the features into a higher-dimensional space when the features are not linearly separable. In addition, Mag-LBP provided the highest ACC of 93.94%, and the SEN and SPE are also good. In the case of ConcSM-LBP, the best obtained ACC is 97.73%, and the AUC is 0.98, which is very close to 1 and

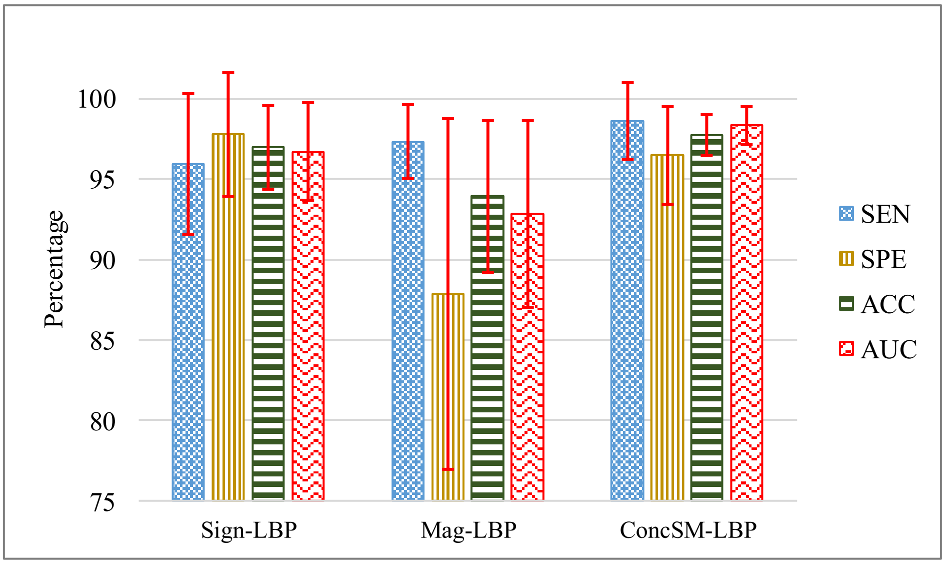

FIGURE 7. A comparison of performance measures with feature selection.

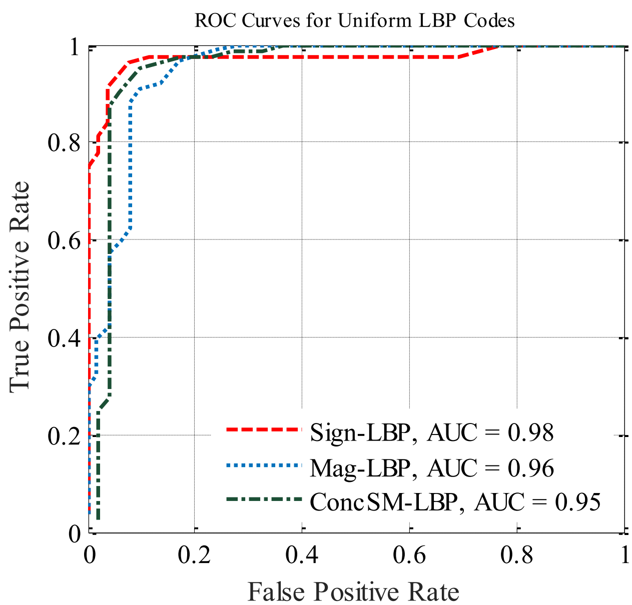

FIGURE 8. ROC curves for Sign-LBP, Mag-LBP, and ConcSM-LBP for best classification accuracy.

shows that the obtained results of the proposed software are reliable.

A comparison of the performance measures SEN, SPE, ACC, and AUC for Sign-, Mag-, and ConcSM-LBP for the case of feature selection is provided in Fig. 7. The maximum obtained ACC is 97.73% with an error of 1.2 for ConcSM-LBP with the top three features. In addition, the best ACC for Sign-LBP is 96.97% and for Mag-LBP it is 93.94%.

IV. EXPLORATION AND DEPLOYMENT OF THE PROPOSED SOFTWARE

The working of the proposed software and how it discriminates between normal and disordered speech signals is described in this section. In addition, deployment of the proposed software is also discussed.

A. EXPLORATION OF THE PROPOSED SOFTWARE

Two types of LBP codes, Sign-LBP and Mag-LBP, are computed from the recurrence plots in the proposed healthcare software. The histograms of LBP codes contained 256 bins, and they are reduced to 59 bins by applying uniform mapping on Sign-LBP and Mag-LBP. The obtained 59 bins (including58 for uniform LBP codes and one for all uniform codes) are treated as features and used to screen the subjects. The evaluation of the software is performed with histograms of uniform Sign-LBP and Mag-LBP, one by one. The highest obtained ACC for uniform Sign-LBP is 96.21% $\pm$ 1.3, and AUC is 0.98, whereas for uniform Mag-LBP, the obtained maximum ACC is 95.45% $\pm$ 2.3 with AUC equal to 0.96. In addition, both histograms are also concatenated (ConcSM-LBP) for the screening of subjects. The obtained maximum ACC for uniform ConcSM-LBP is the same as for uniform Sign-LBP, although the AUC is smaller potentially because of the dimensions of the histograms. The number of bins for ConcSM-LBP is double than Sign-LBP (i.e. 118 bins), and therefore the features are not linearly separable and also not reliable as AUC degraded from 0.98 to 0.95. The ROC curves for each of the best cases are presented in Fig. 8.

However, an ACC of 96.21% is obtained with the proposed method, which is really good, especially when existing screening software [34] are unable to perform well for Type 2 and Type 3 speech signals. However, the number of features is 59 for Sign-LBP and 118 for ConcSM-LBP, which is definitely a large quantity. Therefore, features are sorted according to their FDR so that the most discriminant features can be determined. For each case (Sign-LBP, Mag-LBP, and ConcSM-LBP), the top three features are selected

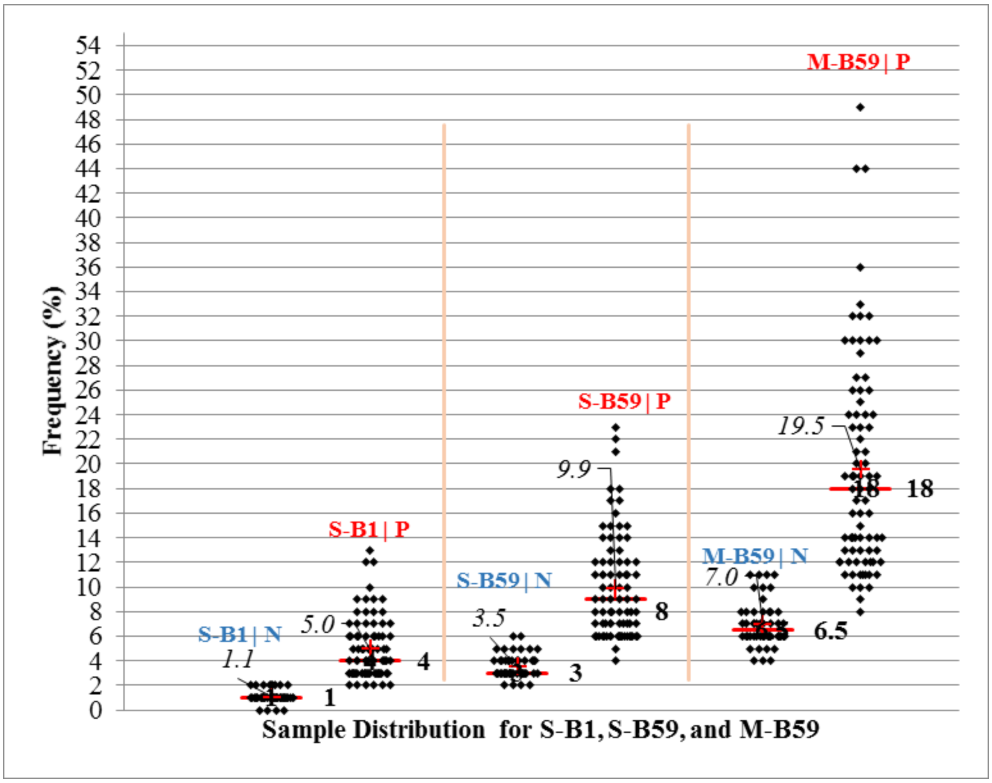

FIGURE 9. Scattergrams for the top three features: S-B1, S-B59, and M-B59.

TABLE 3. Statistical analysis for the top three features: S-B1, S-B59, and M-B59.

Statistics	S-B1 | N	S-B1 | P	S-B59 | N	S-B59 | P	M-B59 | N	M-B59 | P
Minimum	0%	2%	2%	4%	4%	8%
Maximum	2%	13%	6%	23%	11%	49%
Freq. of minimum	4	6	4	1	3	1
Freq. of maximum	9	1	2	1	4	1
Range	2%	11%	4%	19%	7%	41%
1st Quartile	1%	3%	3%	6%	6%	13%
Median	1.0%	4%	3%	9%	6.5%	18%
3rd Quartile	1	6%	4%	12%	8%	24%
Mean	1.09%	4.96%	3.51%	9.89%	6.96%	19.53%
STD	0.49%	2.43%	0.93%	4.15%	1.77%	8.59%

for evaluation of the proposed healthcare software. With the top three features, the maximum ACC for Sign-LBP and Mag-LBP is 96.97% $\pm$ 2.6 and 93.94% $\pm$ 1.3, respectively. The classification ACC for ConcSM-LBP with the top three features is 97.73% $\pm$ 1.2.

In the studies [16] and [18] conducted by Arjmandi et al. and Al-Nasheri et al., respectively, the ACC for screening of normal and disordered subjects for the MEEI subset is 89.3% in [16] and 89.7% in [18]. The reason for the lower classification ACC is the inclusion of Type 2 and Type 3 signals in the experiments, because traditional acoustic features provided by the Multi-Dimensional Voice Program (MDVP) [34] are used, and they are not good for non-periodic signal. The proposed software in this study provided an ACC of 97.73%, which is really good. Hence, it can be used with non-periodic signals of Type 2 and Type 3 to achieve high accuracy.

It can be observed from Tables 1 and 2 that the best obtained ACC is 97.73% with a minimum error of $\pm$1.2, and this result is obtained by the top three features of ConcSM-LBP. These three features are bin 1 of Sign-LBP (S-B1), bin 59 of Sign-LBP (S-B59), and bin 59 of Mag-LBP (M-B59). In every histogram, bin 1 represents the frequency of zero in a recurrence plot, and it shows that there is no change in the neighbors with respect to the center element in a recurrence plot. In addition, the irregularities in the vibrations of the vocal folds make the voice weaker, whispery, and breathier, and hence the speech sample becomes more transient. Due to the transient nature of the disordered speech

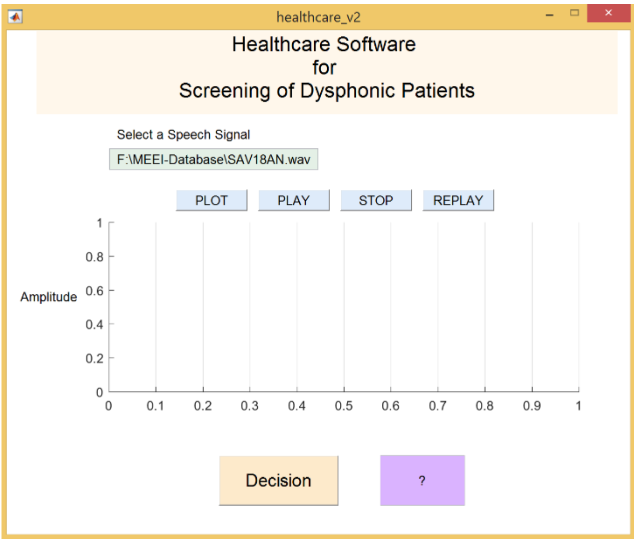

FIGURE 10. **The GUI of the proposed healthcare software.**

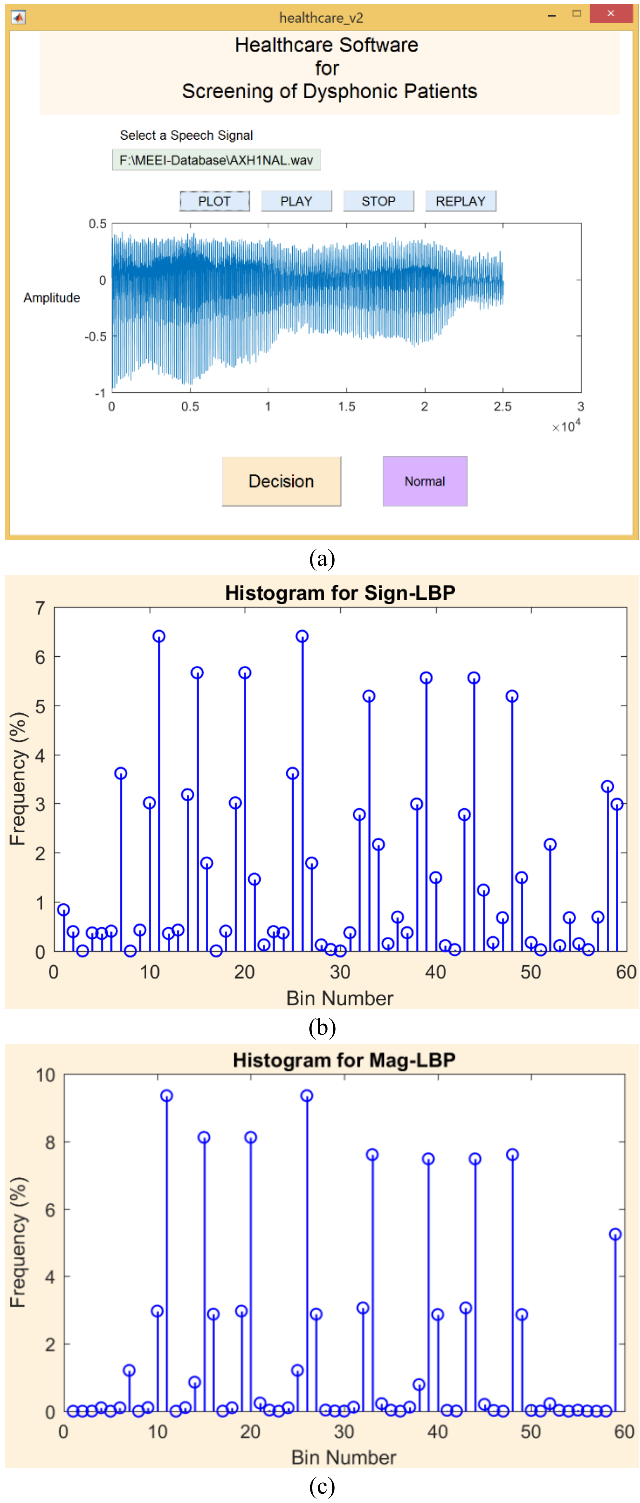

FIGURE 11. **(a) Evaluation of a normal speech signal, (b) Sign histogram of the LBP codes for a normal signal, and (c) Magnitude histogram of the LBP codes for a normal signal.**

sample, the number of non-uniform LBP codes is larger than that of the normal subjects. Bin 59 of Sign-LBP and Mag-LBP represents the frequency of non-uniform LBP codes in Sign- and Mag-LBP histograms.

The contribution of the top three features of ConcSM-LBP is observed by performing statistical analysis. As mention in Fig. 2, the dimension of the generated recurrence plot is 512×512. As no zero-padding is done, the number of LBP codes in a histogram is $510 * 510 = 2{,}60{,}100$. During statistical analysis, for the sake of simplicity, the frequency of LBP codes in histogram bins are represented by the percentage. For example, if the frequency of zero in bin 1 is 2601, then it will be represented by 1%. Scattergrams for all normal and disordered subjects for the top three features are depicted in Fig. 9, where the mean values are highlighted in bold and the median values are in italics. The statistics of the top three features are listed in Table 3.

In Fig. 9, It can be observed that the mean and median of S-B1, bin 1 of Sign-LBP, for normal and disordered subjects is (1.09, 1) and (4.96, 5), respectively. There is a significant difference between the means and medians of normal and disordered subjects. A two-tail t-test is performed to get the p-value at the 5% significance level. The obtained p-value is 0.1E-10 (<0.05), and it rejects the null hypothesis that the difference between the means of normal and disordered subjects for S-B1 is zero. For the second feature, S-B59 (bin 59 for Sign-LBP), the mean and median for normal and disordered subjects are (3.51, 3) and (9.98, 9), respectively. The large difference between means and medians suggests that disordered speech signals are more transient. Therefore, non-uniform codes are very large, and thus around 3% of the codes are non-uniform in normal histograms and 10% of the codes are non-uniform for disordered histograms. Similarly, a large difference is observed for M-B59, bin 59 for Mag-LBP. The means and medians for this case are (6.96, 6.5) and (19.53, 18) for normal and disordered subjects, respectively. Non-uniform codes in a disordered histogram are approximately 20%, which shows the transient nature of the disordered subjects.

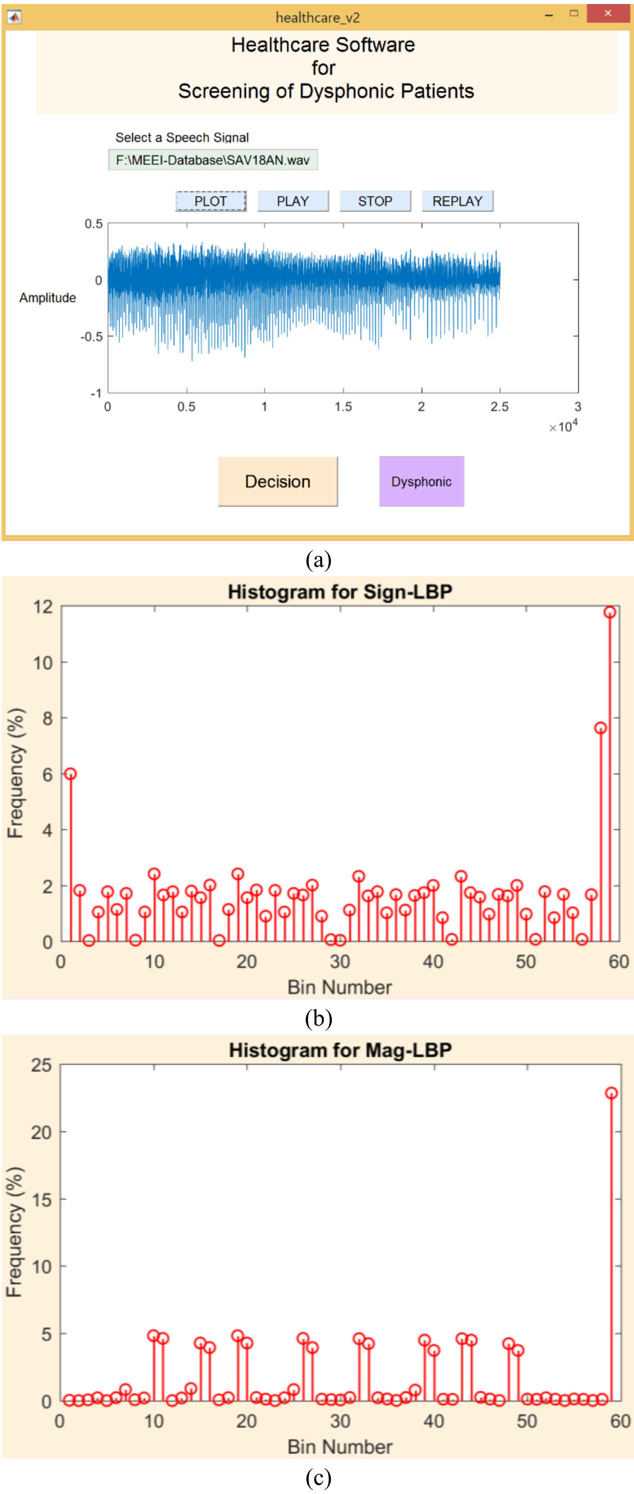

FIGURE 12. (a) Evaluation of a disordered speech signal, (b) Sign histogram of the LBP codes for a disordered signal, (c) and Magnitude histogram of the LBP codes for a disordered signal.

B. DEPLOYMENT OF THE PROPOSED SOFTWARE

The Graphical User Interface (GUI) of the developed healthcare software is depicted in Fig. 10, which shows the main screen of the software. To evaluate a signal for the detection of a voice disorder, a user can select a speech signal via the main screen. The proposed software provides different options to manipulate the selected speech signal, and thus a user can plot, play, stop, and replay it.

After selecting a speech signal, a user will click on the decision button located on the main screen to evaluate the signal. The decision button will provide the judgement on the selected signal, revealing whether it belongs to a normal person or a dysphonic patient. As shown in Fig. 11(a), a user has selected a signal that the software has subsequently evaluated as that of a normal person who does not have any voice complications. With the decision, the proposed software also provides justification by plotting the sign and magnitude histograms (Sign-LBP and Mag-LBP), which are shown in Figs. 11(b) and (c), respectively. As mentioned in Table 3, that mean value of S- B1 for a normal person is about 1% and it can be verified in Fig. 11(b) that the value of the first bin (S-B1) is the same. Furthermore, S-B59 and M-B59 in Figs. 11(b) and 11(c), respectively, are similar to those provided in Table 3 for a normal person.

In Fig. 12(a), a user has selected a disordered signal and the decision of the proposed software is that it belongs to a dysphonic patient. Three bins (S-B1, S-B59, and MB59) provide a clear clue that the person is suffering from a voice disorder because the values of the bins are similar to those mentioned in Table 3 for disordered persons. The proposed software can be deployed in clinics to provide complementary information to the medical doctors by ensuring an accurate diagnosis of voice disorders. In particular, the proposed software can be deployed in remote areas where no specialized doctors are available. In such areas, general physicians can use this software to evaluate people and can refer a person to specialized clinics in the case of any voice complication.

V. CONCLUSION

A healthcare software program for screening dysphonic patients with non-periodic signals is designed and implemented in this study, which is a challenging task due to the presence of noisy components and strong variation of F0 in non-periodic signals. The proposed software does not rely on the computation of F0 because an accurate estimation of F0 is one of the major risks of software failure. Existing software utilize traditional acoustic features that are strongly dependent on F0. Therefore, the possibility of incorrect diagnosis is very high, which in some circumstances can be life threatening for patients. This is the reason that a new type of feature is used in the proposed healthcare software, in order to avoid any type of software failure risk. The features in the proposed software are computed by using recurrence plots and an LBP operator. The proposed software is tested and evaluated by using the MEEI voice disorder database, and the maximum obtained ACC is 96.21%±1.3 by using the full set of features. However, the ACC of the proposed software with selection is increased to 97.73% ± 1.2. The area under the ROC curve closest to 1 suggests that the decisions of the proposed software are reliable. The proposed software can be deployed in the clinics to facilitate the medical doctors in accurate decision making regarding the screening of dysphonic patients. In addition, the proposed software can be used in web-based applications to avoid hospital visits and long waiting time in clinics.

REFERENCES

[1] M. W. Bovee, D. L. Paul, and K. M. Nelson, "A framework for assessing the use of third-party software quality assurance standards to meet FDA medical device software process control guideline's," *IEEE Trans. Eng. Manag.*, vol. 48, no. 4, pp. 465–478, Nov. 2001.

[2] F. McCaffery, D. McFall, P. Donnelly, F. G. Wilkie, and R. Sterritt, "A software process improvement lifecycle framework for the medical device industry," in *Proc. 12th IEEE Int. Conf. Workshops Eng. Comput.-Based Syst. (ECBS)*, Apr. 2005, pp. 273–280.

[3] K. R. Linberg, "Defining the role of software quality assurance in a medical device company," in *Proc. 6th Annu. IEEE Symp. Comput.-Based Med. Syst.*, Jun. 1993, pp. 278–283.

[4] D. R. Wallace and D. R. Kuhn, "Lessons from 342 medical device failures," in *Proc. 4th IEEE Int. Symp. High-Assurance Syst. Eng.*, Nov. 1999, pp. 123–131.

[5] D. R. Wallace and D. R. Kuhn, "Failure modes in medical device software: An analysis of 15 years of recall data," *Int. J. Rel., Quality Safety Eng.*, vol. 8, no. 4, pp. 351–371, 2001.

[6] I. Lee *et al.*, "High-confidence medical device software and systems," *Computer*, vol. 39, no. 4, pp. 33–38, Apr. 2006.

[7] J. Armour and W. S. Humphrey, "Software product liability," Tech. Rep. CMU/SEI-93-TR-1, ESC-TR-93-190, 1993.

[8] R. Rakitin, "Coping with defective software in medical devices," *Computer*, vol. 39, no. 4, pp. 40–45, Apr. 2006.

[9] W. S. Richard, "Software risk management," in *Software Engineering: Barry W. Boehm, Lifetime Contributions to Software Development, Management, and Research*. Hoboken, NJ, USA: Wiley, 2007, pp. 383–497.

[10] B. W. Boehm, "Software risk management: Principles and practices," *IEEE Softw.*, vol. 8, no. 1, pp. 32–41, Jan. 1991.

[11] R. Jardim, S. M. Barreto, and A. Assunção, "Voice Disorder: Case definition and prevalence in teachers," *Revista Brasileira Epidemiologia*, vol. 10, no. 4, pp. 625–636, 2007.

[12] N. Roy, R. M. Merrill, S. D. Gray, and E. M. Smith, "Voice disorders in the general population: Prevalence, risk factors, and occupational impact," *Laryngoscope*, vol. 115, no. 11, pp. 1988–1995, 2005.

[13] N. Roy, R. M. Merrill, S. Thibeault, R. A. Parsa, S. D. Gray, and E. M. Smith, "Prevalence of voice disorders in teachers and the general population," *J. Speech Lang Hear Res.*, vol. 47, pp. 93–281, Apr. 2004.

[14] I. Titze, *Workshop on Acoustic Voice Analysis: Summary Statement*, National Center for Voice and Speech, Denver, CO, USA, 1995.

[15] M. Brockmann, M. J. Drinnan, C. Storck, and P. N. Carding, "Reliable jitter and shimmer measurements in voice clinics: The relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task," *J. Voice*, vol. 25, no. 1, pp. 44–53, Jan. 2011.

[16] M. K. Arjmandi, M. Pooyan, M. Mikaili, M. Vali, and A. Moqarehzadeh, "Identification of voice disorders using long-time features and support vector machine with different feature reduction methods," *J. Voice*, vol. 25, no. 6, pp. e275–e289, Nov. 2011.

[17] B. Boyanov and S. Hadjitodorov, "Acoustic analysis of pathological voices. A voice analysis system for the screening of laryngeal diseases," *IEEE Eng. Med. Biol. Mag.*, vol. 16, no. 4, pp. 74–82, Jul. 1997.

[18] A. Al-Nasheri *et al.*, "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. 2016, doi: 10.1016/j.jvoice.2016.03.019.

[19] J. J. Jiang, Y. Zhang, J. MacCallum, A. Sprecher, and L. Zhou, "Objective acoustic analysis of pathological voices from patients with vocal nodules and polyps," *Folia Phoniatrica Logopaedica*, vol. 61, no. 6, pp. 342–349, 2009.

[20] J. P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *EPL (Europhysics Letters)*, vol. 4, no. 9, p. 973, 1987.

[21] N. Marwan, "How to avoid potential pitfalls in recurrence plot based data analysis," *Int. J. Bifurcation Chaos*, vol. 21, no. 4, pp. 1003–1017, 2011.

[22] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Phys. Rep.*, vol. 438, nos. 5–6, pp. 237–329, 2007.

[23] D. Angus, A. Smith, and J. Wiles, "Conceptual recurrence plots: Revealing patterns in human discourse," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 6, pp. 988–997, Jun. 2012.

[24] P. L. Dhingra and S. Dhingra, *Diseases of Ear, Nose and Throat*, 6th ed. New York, NY, USA: Elsevier, 2014.

[25] J. Kreiman, B. R. Gerratt, G. B. Kempster, A. Erman, and G. S. Berke, "Perceptual evaluation of voice qualityreview, tutorial, and a framework for future research," *J. Speech, Lang., Hearing Res.*, vol. 36, no. 1, pp. 21–40, 1993.

[26] B. R. Gerratt, J. Kreiman, N. Antonanzas-Barroso, and G. S. Berke, "Comparing internal and external standards in voice quality judgments," *J. Speech, Lang., Hearing Res.*, vol. 36, no. 1, pp. 14–20, 1993.

[27] J. L. Sofranko and R. A. Prosek, "The effect of experience on classification of voice quality," *J. Voice*, vol. 26, no. 3, pp. 299–303, 2012.

[28] J. Kreiman and B. R. Gerratt, "Sources of listener disagreement in voice quality assessment," *J. Acoust. Soc. Amer.*, vol. 108, no. 4, pp. 1867–1876, 2000.

[29] T. Mau, "Diagnostic evaluation and management of hoarseness," *Med. Clin. North Amer.*, vol. 94, no. 5, pp. 945–960, Sep. 2010.

[30] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[31] O. Lahdenoja, J. Poikonen, and M. Laiho, "Towards understanding the formation of uniform local binary patterns," *ISRN Mach. Vis.*, vol. 2013, p. 20, Jun. 2013.

[32] C. Shan and T. Gritti, "Learning discriminative LBP-histogram bins for facial expression recognition," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, London, U.K., 2008, Art. no. 429347.

[33] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[34] *Muti-Dimensional Voice Program (MDVP) Ver. 3.3*, Kay Elemetric Corp., Lincoln Park, NJ, USA, 1993.

[35] Massachusette Eye & Ear Infirmry Voice & Speech LAB, *Disordered Voice Database Model 4337 (Ver. 1.03)*, Lincoln Park, NJ, USA: Kay Elemetrics Corp., 1994.

[36] T. Villa-Cañas *et al.*, "Automatic detection of laryngeal pathologies using cepstral analysis in Mel and Bark scales," in *Proc. Symp. Image, Signal Process., Artif. Vis. (STSIVA)*, Sep. 2012, pp. 116–121.

[37] J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechión, V. Osma-Ruiz, G. Castellanos-Domínguez, "An improved method for voice pathology detection by means of a HMM-based feature space transformation," *Pattern Recognit.*, vol. 43, no. 9, pp. 3100–3112, 2010.

[38] G. Muhammad and M. Melhem, "Pathological voice detection and binary classification using MPEG-7 audio features," *Biomed. Signal Process. Control*, vol. 11, pp. 1–9, May 2014.

[39] M. Markaki and Y. Stylianou, "Voice Pathology detection and discrimination based on modulation spectral features," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 7, pp. 1938–1948, Sep. 2011.

[40] J. I. Godino-Llorente, P. Gomez-Vilda, and M. Blanco-Velasco, "Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 10, pp. 1943–1953, Oct. 2006.

[41] V. Parsa and D. G. Jamieson, "Identification of pathological voices using glottal noise measures," *J. Speech, Lang., Hearing Res.*, vol. 43, pp. 469–485, Apr. 2000.

[42] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. Workshop Comput. Learn. Theory*, 1992, pp. 144–152.

[43] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[44] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

ZULFIQAR ALI received the master's degree in computational mathematics from the University of the Punjab, Lahore, and the master's degree in computer science from the University of Engineering and Technology, Lahore, with the specialization in system engineering. Since 2010, he has been a full-time Researcher with the Digital Speech Processing Group, Department of Computer Engineering, King Saud University, Saudi Arabia. He is also a member of the Center for Intelligent Signal and Imaging Research, Universiti Teknologi PETRONAS, Malaysia. His current research interests include speech and language processing, medical signal processing, privacy and security in healthcare, multimedia forensics, and computer-aided pronunciation training systems.

MUHAMMAD TALHA received the Ph.D. degree in computer science from the Faculty of Computing, University of Technology, Malaysia. He is currently with the Deanship of Scientific Research, King Saud University, Riyadh, Saudi Arabia. His research interests include image processing, medical imaging, features extraction, and classification and machine learning techniques.

MANSOUR ALSULAIMAN received the Ph.D. degree from Iowa State University, USA, in 1987. Since 1988, he has been associated with the Computer Engineering Department, King Saud University, Riyadh, Saudi Arabia, where he is currently a Professor with the Department of Computer Engineering. His current research interests include automatic speech/speaker recognition, automatic voice pathology assessment systems, computer-aided pronunciation training system, and robotics. He was an Editor-in-Chief of the *King Saud University Journal Computer and Information Systems*.