Received April 9, 2019, accepted April 17, 2019, date of publication April 26, 2019, date of current version May 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2913444

Dysphonia Detection Index (DDI): A New Multi-Parametric Marker to Evaluate Voice Quality

LAURA VERDE[1], GIUSEPPE DE PIETRO[1], (Member, IEEE), MUBARAK ALRASHOUD[2], (Senior Member, IEEE), AHMED GHONEIM[2], (Member, IEEE), KHALED N. AL-MUTIB[2], AND GIOVANNA SANNINO[1], (Member, IEEE)

[1]Institute of High Performance Computing and Networking (ICAR-CNR), 111 Naples, Italy
[2]Department of Software Engineering (SWE), College of Computer and Information Sciences (CCIS), King Saud University, Riyadh 11543, Saudi Arabia

Corresponding author: Laura Verde (laura.verde@icar.cnr.it)

The authors extend their appreciation to the Deanship of Scientific Research at King Saud University, Riyadh, Saudi Arabia for funding this work through the research group project no. RG-1440-135.

ABSTRACT The rapid diffusion of voice disorders and the lack of appropriate knowledge about the problem have prompted the search for novel and reliable approaches to detect dysphonia, through easy and accessible instruments such as mobile devices. These systems represent, in fact, valid instruments to improve the patient care not only to facilitate the monitoring of symptoms of any diseases but also supporting the correct diagnosis of pathology, such as the dysphonia. In this paper, we propose a new marker, namely the dysphonia detection index, able to support the evaluation of voice disorders, which can be embedded in a mobile health solution. Four acoustic parameters are combined in a single marker to globally evaluate the state of the health of the voice and to assess the presence or not of a voice disorder. A model tree regression algorithm has been applied to define the relationship between these parameters, and the Youden analysis has been used to define the threshold value to distinguish a pathological from a healthy voice. The reliability of the proposed index has been tested in terms of correct classification of accuracy, sensitivity, and specificity. A dataset of 2003 voices has been used to evaluate the performance of our proposed index, composed of samples selected from three different databases: the Massachusetts Eye and Ear Infirmary, the Saarbruecken Voice, and the VOice ICar fEDerico II databases. Our approach achieved the best performances in comparison with other algorithms, and accuracy equals to 82.2%, while sensitivity and specificity are 82% and 82.6%, respectively.

INDEX TERMS Voice diseases, signal processing, mobile health system, classification accuracy, voice assessment.

I. INTRODUCTION

Mobile health (m-health) systems are currently assuming an increasingly important role in the assessment of the state of health. Several m-health systems can deliver healthcare anytime and anywhere providing support for the monitoring, treatment and diagnosis of a specific disease. These systems constitute an helpful instrument to support clinical decisions and conduct research to improve patient care, such as the solutions described in [1]–[3] A considerable amount of heterogeneous data can be acquired and collected by a patient using her/his mobile device, such as a smartphone or tablet. The variety and complexity of these data require the provision of new models, technologies and tools able to process and analyses them in a reliable and easy way. Artificial Intelligence (AI) algorithms offer the opportunity to achieve this objective [4]. There is a currently a wide interest in the adoption of AI techniques in relation to healthcare applications just as has been demonstrated in various areas of science and engineering. AI techniques are applied for medical

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

imaging [5], [6] or signal processing [7]–[9]. These offer an opportunity to improve the monitoring and detection of specific diseases, such as dysphonia.

Dysphonia is an alteration of the voice quality due to morphological and functional alterations of the pneumo-phono-articulatory apparatus. The diagnosis of voice disorders presently requires several medical examinations, as indicated by the SIFEL (Societá Italiana di Foniatria e Logopedia - the Italian Society of Logopedics and Phoniatrics) protocol [10]. Some of these examinations are invasive and must necessarily be performed by medical expert specialists, such as for example the laryngoscopy. Other examinations, instead, are performed by using appropriate tools, such as the acoustic analysis. This consists of the estimation and evaluation of several specific parameters extracted by sustained vowel phonations.

The Fundamental Frequency (F_0), jitter, shimmer and Harmonic to Noise Ratio (HNR) constitute the main acoustic parameters evaluated in clinical practice. Their estimations are often performed by opportune tools, such as the Multi-Dimensional Voice Program (MDVP) [11] or Praat [12] (its name deriving from the imperative form of ''praaten'', ''to speak'' in Dutch), two software systems able to estimate these parameters but not analyze them. These estimations, in fact, can be interpreted only by a medical specialist who is able to indicate the presence of possible laryngeal alterations. Other systems, instead, such as those described in [13]–[15], estimate the acoustic parameters and provide their interpretation, indicating to the user the presence or absence of possible alterations. Unfortunately, these systems are limited to an evaluation of the acoustic parameters individually and do not offer a global measure of vocal quality. Additionally, the presence of voice disorders is estimated comparing the obtained value with a fixed healthy range. The voice is healthy if the value of estimated parameter is included in this healthy range, pathological otherwise. The choice of appropriate healthy range is a crucial point, because there is not a standard healthy range. This can, in fact, change from laboratory to laboratory, influencing scale validity and reliability.

In this study, we propose a new marker, the Dysphonia Detection Index (DDI), that evaluates globally the voice quality through a multi-parametric approach. The main acoustic parameters have been considered, such as the F_0, jitter, shimmer and HNR. An opportune regression algorithm has been used to automatically find a relationship between these parameters, while the Youden analysis has been employed to find the threshold value to estimate the presence or not of a voice disorder. The performance of the DDI has been tested on a wide dataset composed of voices selected from three different databases.

This paper is organized as follows. In Section II, the main studies relating to the search for a valid index able to distinguish between healthy and pathological voices are presented. Section III introduces the proposed approach, while the experimental phase and the achieved results are discussed in Section IV. Finally, our conclusions are presented in Section V.

II. RELATED WORK

Several rating scales and systems have been proposed in literature to assess voice disorders, as for example the GRBAS and the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) scales [16]. The first includes five qualitative characteristics: Grade of dysphony (G), Roughness (R), Breathiness (B), Asthenia (A) and Strain (S). A value between 0 and 3 is assigned to each of these characteristics, where 0 corresponds to a healthy voice, and 1 to light, 2 to moderate and 3 to severe disease. Like the GRBAS scale, the CAPE-V consists in assigning a rating of severity (i.e. the grade of the GRBAS scale), roughness, breathiness and strain. The ranges of the four points of the GRBAS scale are substituted by a visual analogue scale (VAS), on which visual markers for mild, moderate and severe ratings are placed. Although these scales are easy to use and their reliability has been demonstrated in several works such as [17], [18], it is important to note that these evaluations must be performed by a specialist. They, in fact, represent a subjective assessment based on the perception and expertise of the evaluator.

These considerations have encouraged clinicians and researchers to develop methods to evaluate objectively the voice quality by using, in some cases, the parameters evaluated by the acoustic analysis, obtained by processing the voice signal through opportune methodologies. The Dysphonia Severity Index (DSI) [19] is an example. It evaluates the vocal quality basing on a weighted combination of a set of measurements, such as the highest frequency (F_0-high), lowest intensity (I-Low), maximum phonation time (MPT) and jitter. Their linear combination, calculated by logistic regression, reflects the degree of hoarseness, expressed by the G of the GRBAS scale. The reliability of the DSI has been estimated by evaluating the agreement between the observed and predicted perceived voice quality expressed by G. Tested on a private dataset composed of 387 subjects (319 pathological and 68 healthy), a perfect agreement was obtained in 50% of the cases. Additionally, a comparison between the DSI and Voice Handicap Index (VHI) [20] was evaluated, obtaining a high correlation between both measures.

A comparison between CAPE-V and a spectral/cepstral-based acoustic index proposed by Awan et al., is presented in [21]. This index was constructed by using a multiple regression analysis of the Cepstral Peak Prominence (CPP), the spectral ratio and their standard deviations. Tests, performed on a private dataset composed of only 32 subjects (24 pathological and 8 healthy) showed a high correlation with CAPE-V, obtaining a Receiver Operating Characteristic (ROC) area of 0.79, while the sensitivity and specificity were, respectively, 72% and 80%.

Maryn *et al.* [22], instead, proposed the Acoustic Voice Quality Index (AVQI) based on a multiple regression equation constructed by evaluating, as the coefficients, the CPP,

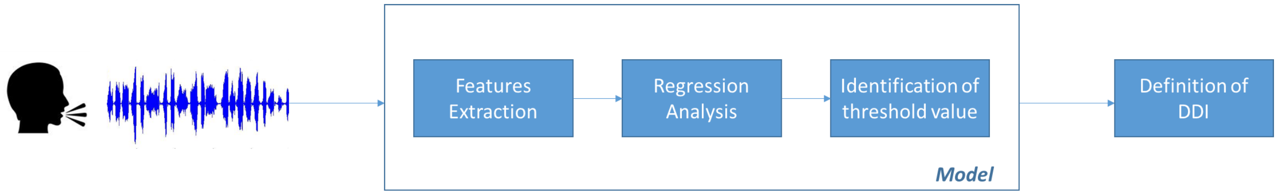

FIGURE 1. Flowchart illustrating the procedure to determine DDI.

HNR, shimmer and spectral measures (slope and tilt). The diagnostic accuracy of the AVQI and its ability to distinguish between healthy and pathological voices was tested by using a private dataset of only 23 vocally normal cases and 228 dysphonic subjects. A ROC-curve was evaluated, which proved to be equal to 0.895. To estimate the sensitivity and specificity, two AVQI cut-off scores, equal to 2.36 and 2.95, were considered. In the first case, the sensitivity and specificity were, respectively, 91% and 59%, while a sensitivity of 74% and specificity of 96%were estimated by using the other AVQI cut-off criteria.

III. THE PROPOSED APPROACH

The DDI marker is able to classify a voice as healthy or pathological by combining information relating to acoustic parameters extracted from voice signals and subjects data such as gender and age.

Figure 1 shows a flowchart illustrating the process employed to estimate the DDI and classify a voice sample as healthy or pathological. Voice signals can be selected from appropriate databases or acquired by using opportune mobile devices, such as a smartphone or tablet. These signals are processed to extract characteristic acoustic features, through opportune techniques or specific m-health solutions, such as [14]. Acoustic features are necessary to build the regression model used to define the DDI index. A Youden analysis is conducted in order to select an appropriate threshold value, necessary to classify a voice sample as healthy or pathological.

A. DATASET

In this study, voice samples extracted by three different databases, the Massachusetts Eye and Ear Infirmary (MEEI) [23], Saarbruecken Voice Database (SVD) [24] and VOice ICar fEDerico II (VOICED) [25] database, were considered. The possibility of testing our approach on samples selected from different databases has enabled us to obtain to have a wider number of voice signals and to evaluate samples with several characteristics and language.

The samples extracted by these databases come from people suffering from several voice disorders and speaking three different languages: American English (MEEI), German (SVD) and Italian (VOICED). This contributes to define the voice quality index independently by subject's language. Some indeces, in fact, are influenced by speech rate or linguistic factors and require a validation in different languages. AVQI, for example, was validated in English, Dutch, German [26], Korean [27] or French [28]. DDI, instead, is independent by language, and this influence is limited also thanks the adoption of sustained vowel, the vocalization of vowel /a/, considered ''language-independent'' and particularly used in clinical voice assessment [29].

MEEI database was developed by the MEEI Voice and Speech Laboratory. It contains several voice recordings of subjects, both healthy and pathological. Pathological subjects suffering from a wide variety of organic, neuralgic, traumatic and psychogenic voice diseases. The sample frequency for the healthy samples is 50 kHz, while that of the pathological samples is 25 kHz or 50 kHz with a 32-bit resolution. In this study, we have selected 53 healthy voices (mean age, 36 ± 8.4 years) and 372 pathological ones (mean age, 47.8 ± 17.7 years). All the samples have a sample frequency equal to 50 kHz and a resolution of 32-bit.

The samples contained in the ***SVD*** were, instead, collected by the Institute of Phonetics of the University of Saarland in collaboration with the Department of Phoniatrics and Ear, Nose and Throat (ENT) at the Caritas clinic St. Theresia in Saarbruecken. It consists of recordings of sustained /a/, /i/ and /u/ vowels, sampled at 50 kHz with a resolution of 16-bit. Samples of subjects suffering from several voice disorders, including functional and organic pathologies, are contained in this database, freely available online [30]. We selected 685 healthy voices (mean age, 27.6 ± 12 years) and 685 pathological ones (mean age, 50.1 ± 15.2 years).

Finally, we have considered all 208 recordings from the ***VOICED*** database, realized by the Institute of High Performance Computing and Networking of the National Research Council of Italy (ICAR- CNR) in collaboration with the Hospital University of Naples Federico II, available on the PhysioNet website [31]. 58 healthy voices (mean age, 37.2 ± 13.4 years) and 150 pathological ones (mean age, 46.7 ± 12.9 years) constitute this database and all of these were added to our dataset. The pathologies are classified as hyperkinetic or hypokinetic dysphonia, or reflux laryngitis. All the samples have a sample frequency equal to 8 kHz and a resolution of 16-bit.

The complete dataset includes 2003 recordings containing the sustained phonation of the vowel sound /a/, as indicated in the SIFEL protocol [10]. In particular, there are 796 healthy voices (mean age, 28.7 ± 12.1 years) and 1207 pathological ones (mean age, 48.7 ± 15.9 years). Further details are

TABLE 1. Details of the voice signals used in this study.

Database	Category	Gender	No	%
MEEI	Healthy	Female	32	1.6
		Male	21	1.0
	Pathological	Female	228	11.4
		Male	144	7.2
SVD	Healthy	Female	428	21.4
		Male	257	12.8
	Pathological	Female	428	21.4
		Male	257	12.8
VOICED	Healthy	Female	37	1.9
		Male	21	1.0
	Pathological	Female	98	4.9
		Male	52	2.6
Total	*Healthy*	*all*	*796*	*60.3*
	Pathological	*all*	*1207*	*39.7*

provided in Table 1, where we have indicated, for each database, the number (No) of the selected samples for each category (healthy or pathological) and gender (female or male), and the percentage (%) calculated on complete dataset.

B. MODEL

DDI index evaluates globally and objectively the voice quality through the estimation of appropriate acoustic parameters. A regression analysis was conducted to find the relationship between these features. Finally, an opportune threshold value to evaluate the presence of a voice disorder was selected. The following sections described in detail the proposed model.

1) FEATURES EXTRACTION

The F$_0$, jitter, shimmer and HNR was used to evaluate the voice quality. These are the main parameters estimated in the clinical acoustic analysis, able to assess objectively the voice and conditions of pneumo-phono-articulatory apparatus [10]. Each parameter, in fact, represents a specific characteristic of this apparatus. F$_0$, for example, represents the rate of vibration of vocal folds and information about the laryngeal function. While the instabilities of the oscillating of the vocal folds are evaluated by jitter and shimmer. Finally, the noise due to incomplete vocal fold closure, a disorder typical of voice pathologies, is indicated by HNR.

There are no standard algorithms available to estimate the acoustic parameters. In this study, we used our personalized methodology described in [32] to estimate the F$_0$, whose performances are compared with the main algorithms existing in literature, such as the Sawtooth Waveform Inspired Pitch Estimator (SWIPE) [33], Subharmonic-to-Harmonic Ratio Procedure (SHRP) [34], YIN algorithm [35]. The cycle-to-cycle instabilities in frequency and amplitude, respectively the jitter and shimmer, instead, were calculated according to the methods indicated in [36]. In detail, the jitter was estimated as a percentage calculated as the average absolute difference between consecutive periods, divided by the average period. The shimmer, instead, was calculated as the average absolute base-10 logarithm of the difference between the amplitudes of consecutive periods multiplied by 20 and expressed in decibels. Finally, the HNR was estimated with the de Krom's algorithm [37] and expressed in dB.

Additionally, physiological information about user have been used, such as gender and age due to the influence of these characteristics and the progression of voice pathologies [38].

2) REGRESSION ANALYSIS

The DDI index was constructed by using a M5-Pruned (M5P) model tree algorithm [39]. The realization of a model tree is based on a combination between a conventional decision tree and linear regression functions at the leaves, creating a clear decision structure. In detail, M5P is a system able to build tree-based models where regression trees have multivariate linear models at their leaves. Tree-based models are constructed by the divide-and-conquer method: the training set is associated with a leaf or some test is chosen that splits the training set into subsets corresponding to the test outcomes; this process is then applied recursively to the subsets. Every potential test is evaluated by determining the subset of cases associated with each outcome.

It is important to note that the choice of the appropriate algorithm can influence the definition of the DDI index and consequent voice assessment. An experimental study was conducted to select the more reliable model, presented in Section IV-B.

3) IDENTIFICATION OF THRESHOLD VALUE

To evaluate if the obtained DDI value can indicate a possible voice alteration, an appropriate threshold value was selected by using the Youden index (J). This is one of the main summary statistics of the ROC curve, which is commonly used to select the optimal threshold value for a marker, as reported in [40], [41]. The Youden index identifies the optimal demarcation point able to achieve the best balance between sensitivity and specificity.

To evaluate this point, the ROC curve was constructed: the true positive rate (sensitivity) against the false positive rate (1-specificity) were plotted over all possible threshold values (c) of the marker. It is estimated in accordance with the following equation:

$$J_C = max_C \{Sensitivity + Specificity - 1\} \tag{1}$$

The DDI threshold value was equal to 0.753: a voice sample with a DDI value higher than this threshold is considered pathological; otherwise, it is healthy. The analysis was performed using the IBM SPSS Statistics version 25 for Windows.

IV. EXPERIMENTAL PHASE

To evaluate the classification accuracy of the DDI marker in distinguishing a healthy voice from a pathological one, we conducted an appropriate experimental procedure. A comparison was made between the classification accuracy obtained by using our proposed DDI and that obtained from the ordinary rule-based approach by means of the evaluation of acoustic analysis, currently used in clinical practice. Additionally, the performances obtained by the adopted regression model and the main approaches used in literature were compared.

Further details are reported in the following subsections, where we have indicated the performance evaluation metrics used to compare the results achieved by using the ordinary approach used in clinical practice and the main regression algorithms.

A. PERFORMANCE EVALUATION METRICS

To validate and estimate the ability to distinguish between healthy and pathological voices, we evaluated its performances in terms of accuracy, sensibility and specificity. The accuracy, defined as the capability of classifying voice samples correctly, is expressed by Equation 2 as:

$$Accuracy(\%) = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

The sensitivity and specificity indicate, instead, the ability of the model to classify correctly a voice, respectively, as diseased or healthy, and are calculated with Equations 3 and 4:

$$Sensitivity(\%) = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity(\%) = \frac{TN}{TN + FP} \quad (4)$$

where the TP, TN, FP and FN are defined as:

- True Positive (TP): the voice sample is pathological and the marker recognizes this;
- True Negative (TN): the voice sample is healthy and the marker recognizes this;
- False Positive (FP): the voice sample is healthy but the marker recognizes it as pathological; and
- False Negative (FN): the voice sample is pathological but the marker recognizes it as healthy.

To evaluate the classification accuracy of our proposed index, the healthy and pathological samples were randomly divided into training and testing sets. Two different types of tests were performed. Initially, we considered, in fact, a dataset composed of voices selected from all three databases. Subsequently, we adopted an intra-database approach, considering separately the voices selected from the three databases. In this case, both the training and testing sets were composed of samples from the same database. For each database, 70% of the samples constituted the training set, while remaining 30% constituted the testing set. Table 2 shows the distribution of the samples for the training and testing sets considered for each approach.

TABLE 2. The distribution of the samples for training and testing sets.

Set	Category	All databases	Intra-database		
			MEEI	SVD	VOICED
Training	Healthy	561	37	480	44
	Pathological	841	260	479	102
	Total	*1402*	*297*	*959*	*146*
Testing	Healthy	235	16	205	14
	Pathological	366	112	206	48
	Total	*601*	*128*	*411*	*62*

B. COMPARISON WITH THE MAIN REGRESSION ALGORITHMS

A comparison between the adopted regression model and the methods most commonly used in literature was performed to select the more reliable regression model, able to find the best relationship between the considered acoustic parameters in terms of correct classification accuracy. For all the regression models, the Youden index was employed to select the opportune threshold that offered the best balance between sensitivity and specificity and most accurately classified a voice sample as pathological or healthy. The regression models analyzed were:

- *Linear regression (LR)* [42]: this is the most commonly used type of predictive analysis. A linear approach is applied to define the relationships between the data. The threshold value for this model is 0.608;
- *Gaussian Process (GP)* [43]: this computed an weighted average of the noisy observations to reduce them. The regression function is inferred from the given data, reconstructing the underlying signal by removing the contaminating noise. The threshold value is fixed at 0.574;
- *Multi-Layer Perceptron (MLP)* [44]: this is a type of artificial neural network. The general framework of a neural network consists of a three layer architecture constituted by an input layer, one or more hidden layers and the output layer. For our tests, the MLP architecture was composed of 4 hidden layers, with a threshold value equal to 0.649;
- *Simple Linear Regression (SLR)* [45]: this is simple linear regression model, where the attribute that results in the lowest squared error is selected. A result higher than the threshold value equal to 0.565 is an indication of a possible alteration to the vocal tract;
- *Sequential Minimum Optimization for Regression (SMO Reg)* [46]: this implements support vector machine (SVM) for the regression. The improved version was used for the regression, it employs two threshold parameters obtained by evaluating the Karush-Kuhn-Tucker (KKT) conditions. The threshold used to classify a voice as healthy or pathological is fixed at 0.599;
- *Instance-based Learning* [47]: this uses specific instances to achieve the classification predictions. In detail, we used K* that predicts data through an

TABLE 3. Performances of the compared regression algorithms considering the combined databases dataset.

Algorithms	Accuracy (%)	Sensitivity (%)	Specificity (%)
Our Approach	**82.2**	82.0	**82.6**
Linear Regression	79.0	81.4	75.3
Gaussian Process	76.7	83.6	66.0
MultiLayer Perceptron	80.7	86.1	72.3
Simple Linear Regression	76.7	80.1	71.5
SMO Reg	74.2	71.3	78.7
K*	79.0	85.2	69.4
Decision Table	80.2	**97.8**	62.1

TABLE 4. Performance of the compared regression algorithms considering the dataset obtained using intra-database approach.

	MEEI			SVD			VOICED		
Algorithms	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
Our Approach	**98.4**	98.2	**100**	79.8	70.6	90.2	50.0	45.8	64.3
Linear Regression	87.5	**100**	5.8	77.6	64.2	92.7	62.9	66.7	50.0
Gaussian Process	87.5	**100**	5.8	**80.3**	72.5	89.1	69.4	77.1	42.9
MultiLayer Perceptron	**98.4**	98.2	**100**	47.7	0.9	**100**	72.6	87.5	21.4
Simple Linear Regression	87.5	**100**	5.8	77.6	66.5	90.2	67.7	77.1	35.7
SMO Reg	87.5	**100**	5.8	76.4	61.9	92.7	**77.4**	**97.9**	7.1
K*	82.8	80.3	**100**	48.2	2.3	**100**	24.2	2.1	**100**
Decision Table	88.3	**100**	6.25	79.8	**74.3**	86.0	75.8	95.8	7.1

TABLE 5. Healthy range for the acoustic parameters considered.

Parameter	Gender	Healthy range
F_0 (Hz)	Female	189-280 (Hz)
	Male	104-158 (Hz)
Jitter (%)	Female	$\leq$1.04 %
	Male	$\leq$1.04 %
Shimmer (dB)	Female	$\leq$0.35 dB
	Male	$\leq$0.35dB
HNR (dB)	Female	$\leq$20 dB
	Male	$\leq$20 dB

entropy-based distance function, with a threshold value equal to 1;

- *Decision Table (DT)* [48]: this is constituted by two components, a schema, which is a set of features, and a body, consisting of a multiset of labeled instances. An induction algorithm is used to decide the features to include in the schema and the instances to store in the body. The value that offers the best balance between sensitivity and specificity is 0.637.

Table 3 reports a comparison between the results obtained by the regression methods adopted by using the other regression methods existing in literature, considering the combined databases dataset. This table shows that our approach achieved the best performance in terms of accuracy and specificity, respectively equal to 82.2% and 82.6%. The best sensitivity, instead, was achieved by the Decision Table method (97.8%), although its specificity (62.1 %)is lower.

TABLE 6. Performances of our approach and each individual acoustic parameter considering the combined databases dataset.

Algorithms	Accuracy (%)	Sensitivity (%)	Specificity (%)
Our Approach	**82.2**	82.0	**82.6**
F_0 (Hz)	58.2	50.0	71.1
Jitter (%)	52.2	35.2	78.7
Shimmer (dB)	60.9	**92.6**	11.5
HNR (dB)	40.8	25.1	65.1

The results obtained considering the samples selected from the different databases individually are reported in Table 4. Considering the voice samples extracted from the MEEI database, our approach achieved high values of classification accuracy, equal to 98.4%, sensitivity, 98.2% and specificity, 100%. Although the other methods achieved good results in terms of sensitivity, the specificity of our proposed methodology (100%) is higher.

Considering the samples extracted from the SVD database, our approach achieves performances similar to GP, DT or SLR, as shown in Table 3. The accuracy is about 80%, while the sensitivity is 70.6% and specificity 90.2%.

TABLE 7. Performances of our approach and each individual acoustic parameter considering the dataset obtained using intra-database approach.

	MEEI			SVD			VOICED		
Algorithms	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
Our Approach	**98.4**	**98.2**	**100**	**79.8**	70.6	**90.2**	50.0	45.8	**64.3**
F_0 (Hz)	79.7	84.8	43.8	53.5	35.8	73.6	50.0	52.1	42.9
Jitter (%)	43.0	34.8	100	53.0	28.9	80.3	**79.0**	97.9	14.3
Shimmer (dB)	94.5	96.4	81.3	52.3	**93.1**	6.2	67.7	**85.4**	7.1
HNR (dB)	82.8	93.8	6.3	51.8	65.1	36.8	58.1	62.5	42.9

Finally, considering the tests performed on the samples selected from the VOICED database, the best values of accuracy, sensitivity and specificity were not achieved by our approach, but, in comparison with the other regression models, the DDI index offers the best balance between these measurements. K*, for example, achieves the best specificity (100%) but the sensitivity is very low.

C. COMPARISON WITH THE ORDINARY APPROACH

As previously mentioned, currently in clinical practice, the state of health of the voice is evaluated by analyzing and interpreting the acoustic parameters individually, comparing the obtained estimations with a determined healthy ranges, using *if/else* rules, namely:

```
if (estimated value of acoustic parameter is
 within the healthy range){
  Voice classified as healthy
} else {
  Voice classified as pathological
}
```

The difficulty to define an appropriate healthy range can influence the correct estimation of pathology of pneumo-phono-articulatory apparatus. In this study, the healthy ranges considered for the female and male voice samples are indicated in Table 5. The considered ranges to evaluate the jitter, shimmer and HNR are those indicated in [12], [49], [50]. While, the F_0 healthy range was calculated by considering the mean and standard deviation values of the F_0 indicated in these studies [51]–[53].

Table 6 shows a comparison between the results achieved by using our proposed marker and the estimations of each parameter established by the acoustic analysis. The results obtained indicate that, generally, the best performances were achieved using our proposed approach. The classification accuracy achieved, equal to 82.2%, is in fact higher than that obtained by evaluating each individual acoustic parameter. Moreover, although the shimmer obtained a good result in terms of sensitivity (92.6%), the specificity of our approach is higher (82.6% vs 11.5%).

Considering, instead, the tests performed on each database separately, the DDI marker achieved the best performance, in terms of accuracy, in relation to the analyzed samples extracted from the MEEI database, equal to 98.4%, and the best balance between sensitivity and specificity, equal to, respectively, 98.2% and 100%. In this case also, the shimmer achieved a good performance in terms of accuracy, sensitivity and specificity, although these results are lower than those obtained by the DDI, as indicated in Table 7.

Additionally, this table reports the results obtained in evaluating samples selected from the SVD and VOICED databases. In the first case, considering samples extracted from the SVD database, the DDI discriminates between pathological and healthy voices better than each acoustic parameter considered individually. The accuracy is about 80%, sensitivity is about 71% and the specificity 90.2%.

Considering, instead, the performances achieved using voice samples selected from the VOICED database, our approach did not achieve the best classification accuracy, but provided a good balance between sensitivity and specificity. The jitter and shimmer, in fact, achieved a high sensitivity (respectively, 97.9% and 85.4%) but the specificity is low (respectively 14.3% and 7.1%).

V. CONCLUSIONS

The increased use of mobile multimedia services and applications in healthcare offers the opportunity to improve considerably the patient care. These systems, in fact, can help patients to manage their treatments or support better clinical decision making through the integration of opportune techniques of analysis.

In this paper we propose a new multi-parametric acoustic marker able to evaluate globally the voice quality and detect possible disorders, that can be embedded in a mobile health solution, to monitor and support the correct diagnosis of voice disorders. This marker combines, in fact, data provided by each acoustic parameter considered, such as information about the laryngeal function provided by Fundamental Frequency, cycle-to-cycle instabilities in frequency and amplitude by jitter and shimmer, and the presence of noise due to a voice disorder by HNR, to detect any alterations of voice quality due to a possible disorder of the pneumophono-articulatory apparatus.

The proposed approach detects carefully voice disorders. It identifies with a better accuracy (about 82%) the presence of a voice disorder than the approaches required by the standard medical protocols. Additionally, a comparison with other regression algorithms has been performed, which has confirmed the greater accuracy of our approach in detecting a voice disorder than that achieved by using the other algorithms.

As future work, the proposed multi-parametric marker can be embedded into an easy and portable tool able to monitoring the state of voice health. The aim is integrated this marker in a mobile solution, able to evaluate, in real time, globally voice quality by using an easy mobile device, such as a smartphone or tablet.

REFERENCES

[1] M. S. Hossain and G. Muhammad, "Healthcare big data voice pathology assessment framework," *IEEE Access*, vol. 4, pp. 7806–7815, 2016.

[2] G. Muhammad, M. F. Alhamid, M. Alsulaiman, and B. Gupta, "Edge computing with cloud for voice disorder assessment and treatment," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 60–65, Apr. 2018.

[3] Z. Ali, M. S. Hossain, G. Muhammad, and A. K. Sangaiah, "An intelligent healthcare system for detection and classification to discriminate vocal fold disorders," *Future Gener. Comput. Syst.*, vol. 85, pp. 19–28, Aug. 2018.

[4] D. S. Terzi, U. Demirezen, and S. Sagiroglu, "Evaluations of big data processing," *Services Trans. Big Data*, vol. 3, no. 1, pp. 1–11, 2016.

[5] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[6] D. S. Kermany *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

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

[8] G. Sannino and G. De Pietro, "A deep learning approach for ECG-based heartbeat classification for arrhythmia detection," *Future Gener. Comput. Syst.*, vol. 86, pp. 446–455, Sep. 2018.

[9] A. Al-Nasheri *et al.*, "Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions," *IEEE Access*, vol. 6, pp. 6961–6974, 2017.

[10] A. R. Maccarini and E. Lucchini, "La valutazione soggettiva ed oggettiva della disfonia. Il protocollo sifel," *Acta Phoniatrica Latina*, vol. 24, nos. 1–2, pp. 13–42, 2002.

[11] O. Amir, M. Wolf, and N. Amir, "A clinical comparison between MDVP and PRAAT softwares: Is there a difference?" in *Proc. 5th Int. Workshop Models Anal. Vocal Emissions Biomed. Appl.*, 2007, pp. 37–40.

[12] P. Boersma and D. Weenink. (2009). *Praat: Doing Phonetics by Computer (Version 5.1. 05)[Computer Program]*. Accessed: May 1, 2009. [Online]. Available: http://www.fon.hum.uva.nl/praat/

[13] M. Mat Baki *et al.*, "Reliability of OperaVOX against multidimensional voice program (MDVP)," *Clin. Otolaryngol.*, vol. 40, no. 1, pp. 22–28, 2015.

[14] U. Cesari, G. De Pietro, E. Marciano, C. Niri, G. Sannino, and L. Verde, "Voice disorder detection via an m-health system: Design and results of a clinical study to evaluate Vox4Health," *BioMed Res. Int.*, vol. 2018, Aug. 2018, Art. no. 8193694.

[15] E. Van Leer, R. C. Pfister, and X. Zhou, "An iOS-based cepstral peak prominence application: Feasibility for patient practice of resonant voice," *J. Voice*, vol. 31, no. 1, pp. 131-e9–131-e16, 2017.

[16] K. Nemr *et al.*, "GRBAS and cape-V scales: High reliability and consensus when applied at different times," *J. Voice*, vol. 26, no. 6, pp. 812-e17–812-e22, 2012.

[17] S. C. de Almeida, "Validity and reliability of the 2nd European portuguese version of the 'consensus auditory-perceptual evaluation of voice' (II EP CAPE-V)," Ph.D. dissertation, Inst. Politécnico de Setúbal. Escola Superior de Saúde, Setúbal, Portugal, 2016.

[18] R. I. Zraick *et al.*, "Establishing validity of the consensus auditory-perceptual evaluation of voice (CAPE-V)," *Amer. J. Speech-Lang. Pathol.*, vol. 20, no. 1, pp. 14–22, 2011.

[19] F. L. Wuyts *et al.*, "The dysphonia severity index: An objective measure of vocal quality based on a multiparameter approach," *J. Speech, Lang., Hearing Res.*, vol. 43, no. 3, pp. 796–809, 2000.

[20] B. H. Jacobson *et al.*, "The voice handicap index (VHI): Development and validation," *Amer. J. Speech-Lang. Pathol.*, vol. 6, no. 3, pp. 66–70, 1997.

[21] S. N. Awan, N. Roy, M. E. Jetté, G. S. Meltzner, and R. E. Hillman, "Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V," *Clin. Linguistics Phonetics*, vol. 24, no. 9, pp. 742–758, 2010.

[22] Y. Maryn, P. Corthals, P. Van Cauwenberge, N. Roy, and M. De Bodt, "Toward improved ecological validity in the acoustic measurement of overall voice quality: Combining continuous speech and sustained vowels," *J. Voice*, vol. 24, no. 5, pp. 540–555, 2010.

[23] M. Eye and E. Infirmary, *Voice Disorders Database, Version. 1.03 (CD-ROM)*. Lincoln Park, NJ, USA: Kay Elemetrics Corp., 1994.

[24] M. Pützer and J. Koreman, "A german database of patterns of pathological vocal fold vibration," *Phonus*, vol. 3, pp. 143–153, 1997.

[25] U. Cesari, G. De Pietro, E. Marciano, C. Niri, G. Sannino, and L. Verde, "A new database of healthy and pathological voices," *Comput. Elect. Eng.*, vol. 68, pp. 310–321, May 2018.

[26] Y. Maryn, M. De Bodt, B. Barsties, and N. Roy, "The value of the acoustic voice quality index as a measure of dysphonia severity in subjects speaking different languages," *Eur. Arch. Oto-Rhino-Laryngol.*, vol. 271, no. 6, pp. 1609–1619, 2014.

[27] Y. Maryn, H.-T. Kim, and J. Kim, "Auditory-perceptual and acoustic methods in measuring dysphonia severity of Korean speech," *J. Voice*, vol. 30, no. 5, pp. 587–594, 2016.

[28] T. Pommée, Y. Maryn, C. Finck, and D. Morsomme, "Validation of the acoustic voice quality index, version 03.01, in French," *J. Voice*, to be published.

[29] V. Parsa and D. G. Jamieson, "Acoustic discrimination of pathological voice," *J. Speech, Lang., Hearing Res.*, vol. 44, no. 2, pp. 327–339, 2001.

[30] (2018). *Saarbruecken Voice Database*. [Online]. Available: http://http://www.stimmdatenbank.coli.uni-saarland.de/index.php4#target

[31] PhysioNet. (2018). *VOICED (Voice ICar fEDerico II) Database*. [Online]. Available: https://physionet.org/physiobank/database/voiced/

[32] L. Verde, G. De Pietro, and G. Sannino, "A methodology for voice classification based on the personalized fundamental frequency estimation," *Biomed. Signal Process. Control*, vol. 122, no. 5, pp. 2960–2961, 2018.

[33] A. Camacho, *SWIPE: A Sawtooth Waveform Inspired Pitch Estimator for Speech and Music*. Gainesville, FL, USA: Univ. Florida Gainesville, 2007, vol. 42.

[34] X. Sun, "Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 1, May 2002, pp. 1–333.

[35] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Amer.*, vol. 111, no. 4, pp. 1917–1930, 2002.

[36] M. Farrús, J. Hernando, and P. Ejarque, "Jitter and shimmer measurements for speaker recognition," in *Proc. 8th Annu. Conf. Int. Speech Commun. Assoc. (InterSpeech)*, 2007, pp. 778–781.

[37] F. Severin, B. Bozkurt, and T. Dutoit, "HNR extraction in voiced speech, oriented towards voice quality analysis," in *Proc. IEEE 13th Eur. Signal Process. Conf.*, Sep. 2005, pp. 1–4.

[38] R. H. G. Martins, H. A. do Amaral, E. L. M. Tavares, M. G. Martins, T. M. Gonçalves, and N. H. Dias, "Voice disorders: Etiology and diagnosis," *J. Voice*, vol. 30, no. 6, pp. 761.e1–761.e9, Nov. 2016.

[39] J. R. Quinlan, "Learning with continuous classes," in *Proc. 5th Austral. Joint Conf. Artif. Intell.*, vol. 92. Singapore: World Scientific, 1992, pp. 343–348.

[40] R. Fluss, D. Faraggi, and B. Reiser, "Estimation of the youden index and its associated cutoff point," *Biometrical J., J. Math. Methods Biosci.*, vol. 47, no. 4, pp. 458–472, 2005.

[41] M. D. Ruopp, N. J. Perkins, B. W. Whitcomb, and E. F. Schisterman, "Youden index and optimal cut-point estimated from observations affected by a lower limit of detection," *Biometrical J., J. Math. Methods Biosci.*, vol. 50, no. 3, pp. 419–430, 2008.

[42] R. J. Leatherbarrow, "Using linear and non-linear regression to fit biochemical data," *Trends Biochem. Sci.*, vol. 15, no. 12, pp. 455–458, 1990.

[43] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*. New York, NY, USA: Springer, 2003, pp. 63–71.

[44] P. Thomas and M.-C. Suhner, "A new multilayer perceptron pruning algorithm for classification and regression applications," *Neural Process. Lett.*, vol. 42, no. 2, pp. 437–458, 2015.

[45] H. J. Seltman. (2012). *Experimental Design and Analysis*. [Online]. Available: http://www.stat.cmu.edu/ and hseltman/309/Book/Book.pdf

[46] S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to the SMO algorithm for SVM regression," *IEEE Trans. Neural Netw.*, vol. 11, no. 5, pp. 1188–1193, Sep. 2000.

[47] J. G. Cleary and L. E. Trigg, "K*: An instance-based learner using an entropic distance measure," in *Machine Learning Proceedings 1995*. Amsterdam, The Netherlands: Elsevier, 1995, pp. 108–114.

[48] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. San Mateo, CA, USA: Morgan Kaufmann, 2016.

[49] W. De Colle and O. Schindler, *Voce e Computer: Analisi Acustica Digitale del Segnale Verbale (il Sistema CSL-MDVP)*. Italy: Omega, 2001.

[50] I. Guimarães, "A ciência e a arte da voz humana," in *Alcoitão, Escola Superior de Saúde de Alcoitão*. 2007.

[51] G. Williamson, *Human Communication: A Linguistic Introduction*. Milton Keynes, U.K.: Speechmark, 2000.

[52] H. Traunmüller and A. Eriksson, "The frequency range of the voice fundamental in the speech of male and female adults," to be published.

[53] M. P. Gelfer and V. A. Mikos, "The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels," *J. Voice*, vol. 19, no. 4, pp. 544–554, 2005.

LAURA VERDE received the bachelor's degree in biomedical engineering and the master's degree from "Federico II" University, Naples, Italy, in 2009 and 2011, respectively. She is currently pursuing the Ph.D. degree in information engineering with the University of Naples Parthenope. She is also a Research Fellow with the Institute of High Performance Computing and Networking (ICAR), National Research Council of Italy (CNR). Her research interests include biomedical signal processing, voice signal analysis, m-health systems, and artificial intelligence for healthcare.

GIUSEPPE DE PIETRO is currently the Director of the Institute for High Performance Computing and Networking, CNR, and an Adjunct Professor with the College of Science and Technology, Temple University, Philadelphia. He has been actively involved in many European and National projects, with industrial co-operations. He has authored over 200 scientific papers published in international journals and conferences. His current research interests include cognitive computing, clinical decision support systems, and software architectures for e-health. He is an IEEE and KES International Member. He is involved in many program committees and journal editorial boards.

MUBARAK ALRASHOUD received the Ph.D. degree in electrical and computer engineering from Ryerson University, Ottawa, Canada. He is currently an Assistant Professor and the Head of the Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. He is also an Adjunct Professor with the School of Electrical Engineering and Computer Science, University of Ottawa, Canada. He is a Senior Member of the IEEE and ACM.

AHMED GHONEIM (M'10) received the M.Sc. degree in software modeling from the University of Menoufia, Egypt, in 1999, and the Ph.D. degree in software engineering from the University of Magdeburg, Germany, in 2007. He is currently an Associate Professor with the Department of Software Engineering, College of Computer and Information Sciences (CCIS), King Saud University. His research interests include software evolution, service-oriented engineering, software development methodologies, quality of services, net-centric computing, and human–computer interaction (HCI).

KHALED N. AL-MUTIB is currently an Associate Professor with the Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. His research interests include robotics, computation intelligence, and healthcare.

GIOVANNA SANNINO received the bachelor's degree in computer engineering and the master's degree, named European Master on Critical Networked Systems, from the University of Naples Federico II, in 2008 and 2009, respectively, the master's degree *(cum laude)* in telecommunications engineering from the University of Naples Parthenope, in 2011, and the Ph.D. degree in information engineering from the University of Naples "Parthenope," in 2015. She is currently a Researcher with the Institute for the High Performance Computing and Networking (ICAR-CNR) and an Adjunct Professor of informatics with the University of Naples "Federico II." Her research interests include mobile health, pervasive computing, pattern recognition, signal processing, and artificial intelligence for healthcare. In addition, she is an IEEE 11073 Personal Health Device Working Group Member.