

Received March 27, 2020, accepted April 22, 2020, date of publication May 6, 2020, date of current version May 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2992641

Novel Multi Center and Threshold Ternary Pattern Based Method for Disease Detection Method Using Voice

TURKER TUNCER^{ID 1}, SENGUL DOGAN¹, FATİH ÖZYURT^{ID 2}, SAMIR BRAHIM BELHAOUARI^{ID 3}, AND HALIMA BENSMAIL⁴

¹Department of Digital Forensics Engineering, Technology Faculty, Firat University, 23119 Elazığ, Turkey

²Department of Software Engineering, Engineering Faculty, Firat University, 23119 Elazığ, Turkey

³College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

⁴Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar

Corresponding author: Samir Brahim Belhaouari (sbelhaouari@hbku.edu.qa)

ABSTRACT Smart health is one of the most popular and important components of smart cities. It is a relatively new context-aware healthcare paradigm influenced by several fields of expertise, such as medical informatics, communications and electronics, bioengineering, ethics, to name a few. Smart health is used to improve healthcare by providing many services such as patient monitoring, early diagnosis of disease and so on. The artificial neural network (ANN), support vector machine (SVM) and deep learning models, especially the convolutional neural network (CNN), are the most commonly used machine learning approaches where they proved to be performance in most cases. Voice disorders are rapidly spreading especially with the development of medical diagnostic systems, although they are often underestimated. Smart health systems can be an easy and fast support to voice pathology detection. The identification of an algorithm that discriminates between pathological and healthy voices with more accuracy is needed to obtain a smart and precise mobile health system. The main contribution of this paper consists of proposing a multiclass-pathologic voice classification using a novel multileveled textural feature extraction with iterative feature selector. Our approach is a simple and efficient voice-based algorithm in which a multi-center and multi threshold based ternary pattern is used (MCMTTP). A more compact multileveled features are then obtained by sample-based discretization techniques and Neighborhood Component Analysis (NCA) is applied to select features iteratively. These features are finally integrated with MCMTTP to achieve an accurate voice-based features detection. Experimental results of six classifiers with three diagnostic diseases (frontal resection, cordeectomy and spastic dysphonia) show that the fused features are more suitable for describing voice-based disease detection.

INDEX TERMS MCMTTP, discrete wavelet transform, voice disease detection, smart health, machine learning.

I. INTRODUCTION

The voice is one of the most important factors that we use in communication between people. Voice and speaking skills are the easiest way to reflect the thought and it is the most important component that distinguishes it from other living things. The voice is part of the personality and character of almost every person. We can also understand diseases

The associate editor coordinating the review of this manuscript and approving it for publication was Kin Fong Lei^{ID}.

using voices because some diseases directly affected human voice [1]–[3].

Finding diagnoses such as frontal lobe resection, spasmodic dysphonia and cordeectomy from patient voice data has become possible with today's technologies. Little is known about the neuropsychological outcome after frontal resection [2]. However, detection of disease from voice data is still a very new study. Spasmodic dysphonia (SD) is a neurogenic, central originated focal laryngeal dystonia that occurs during speech, with intermittent or continuous spasm of the cord vocals, and muffled sound quality is monitored.

Cordeotomy is the excision of the vocal cord by performing a thyrotomy [3].

Disease detection by using voice has become increasingly popular in the literature over the past few years with the development of medical diagnostic systems [4]. Computer-processed voices are used everywhere these days such as Alexa, Siri, Google Assistant and other services etc. [5], [6].

In traditional medical diagnoses, a medical doctor's report is required, which takes the patient's neurological history and examines various motor skills [7], [8]. It may be necessary to obtain information from many laboratory results for an accurate diagnosis. Therefore, clinical situations that do not lead to misdiagnosis are one of the biggest areas of interest for medical expert systems. In addition, smart medical diagnostic systems have the potential to optimize medical decisions, improve medical treatments, and reduce waiting times and financial costs. Voice recordings have been considered as a potential (noninvasive and low cost) biomarker to diagnose some voice related diseases [9]–[11].

A. MOTIVATION

Voice and sound classification are hot topic research areas of machine learning. In the biomedical engineering applications, automated voice-based disease detection method has been widely used. To develop intelligent health assistant systems, automated voice-based disease detection methods should be developed and enhanced. In the literature, binary classification methods (two class method) have generally been proposed, our main motivation is to perform multiclass-pathologic voice classification by using novel multileveled textural feature extraction and iterative feature selector.

B. RELATED WORKS

This work is about voice signal recognition in biomedical engineering. The main of this work is to reach high accuracy rate by using a naïve method. In this section, some pathological voices classification and detection works were listed in Table 1.

C. OUR METHOD

In this study, we presented a novel one-dimensional TP for feature extraction we called MCMTP as it uses multi centered and multi threshold based ternary pattern. The main objective of the proposed MCMTP is to comprehensively extract features. To generate multileveled features, we use maximum pooling, which is widely used in the deep learning network. The extracted features are selected by using an iterative neighborhood component analysis we called INCA. In the classification phase, 6 classifiers were selected to obtain comprehensive benchmark for evaluating 8 cases.

D. CONTRIBUTIONS

The provided contributions of the proposed multileveled MCMTP based disease classification method are shown in below.

TABLE 1. Selected previously presented voice-based disease detection methods.

Reference	Method	Year	Database	Evaluation Criteria	The results of accuracy
Fujimura et al. method [12]	One-dimensional convolutional neural network	2020	Collected data	Accuracy, F1 score, quadratic weighted cohen's kappa	88.3%
Hammami et al. method [13]	Empirical mode decomposition, discrete wavelet transform	2020	Saarbruecken Voice Database [14]	Accuracy, specificity, sensitivity	99.29
Tracy et al. method [15]	Deep phenotyping method	2019	Collected data	True positive rate, false positive rate, recall, precision, F1, AUC	-
Tuncer et al. method [16]	One dimensional local binary pattern	2019	Saarbruecken Voice Database [14]	Accuracy, specificity, sensitivity, gmean	98.73%
Wu et al. method [17]	Joint learning	2019	Saarbruecken Voice Database [14]	Specificity, sensitivity, gmean	-
Tuncer and Dogan method [18]	Dynamic center based binary ternary	2019	Daily voices dataset [19]	Accuracy	89.17%
Pravenna et al. method [20]	Mel frequency cepstral coefficient, discrete fourier transform	2012	Collected data	Accuracy	98.00%
Haq et al. method [21]	L1-norm support vector machine	2019	Collected data	Accuracy, specificity, sensitivity, precision, F1-score, execution time	97.00%
Bocklet et al. method [22]	Acoustic, vocal and prosodic feature extraction	2011	Rusz et al. dataset [23]	Recognition rate	91.00%
Caesarendra et al. method [24]	Principal component analysis, linear discriminant analysis, support vector machine	2014	California-Irvine University (UCI) machine learning repository [25]	Accuracy	100.0%
Benba et al. method [26]	Human factor cepstral coefficient	2017	Collected data	Accuracy, specificity, sensitivity	87.50%
Naranjo et al. method [2]	Bayesian binary regression	2016	UCI Machine Learning Repository [25, 27]	Accuracy, recall, specificity, precision	87.6%

- 1) TP has been widely used in the literature, but it has some several drawbacks that need to be tackled and enhanced. TP is a parametric feature extractor and it used 5th value as a center pixel. However, more differs and valuable features can be extracted by using variable center and variable threshold. A novel TP like microstructure (MCMTP) is proposed to tackle problems of determining the optimal threshold value, center value selection, and medium-high level

feature extraction. MCMTTP has better feature generating capability than classical TP because of different choices of center values and thresholds to generate features comprehensively as inception network.

- 2) Feature selection is one of the most important steps for machine learning especially, when the number of features is huge. To optimize this phase, an iterative neighborhood component analysis, NCA, feature selector is used. The main aim of the proposed INCA is to automatically select most discriminating features.
- 3) Voice based disease detection is a hard problem for machine learning based methods. Binary classification has generally been used in the voice disease detection method in the literature. Because, multiclass classification is very hard for voice-based disease detection [16], [17], [51]. In this work, frontal resection, cordeectomy and spastic dysphonia diseases are used for performance evaluation and by using these diseases, 8 cases are created. High success rates are reached by using the proposed MCMTTP based multilevelled method.

II. BACKGROUND

In this article, local ternary pattern (LTP) like feature extraction method is used to extract features. Therefore, Local Binary Pattern (LBP) which is basic textural feature extractor and one dimensional LTP is explained in this section.

A. LOCAL BINARY PATTERN

Local Binary Pattern was proposed in 1994 and is a commonly used texture identifier [28]. It is widely used because it is an effective feature extractor [29]. As shown in Figure 1, LBP divides the image into 3×3 sized overlapping blocks. Then the center pixel and neighborhood pixels are compared using the signum function. Mathematical notation of the binary feature extraction is given in Eq. 1 [30], [31].

$$\text{signum}(IM_{t,k}, \text{center}) = bit_t = \begin{cases} 0, & IM_{t,k} < \text{center} \\ 1, & IM_{t,k} \geq \text{center} \end{cases} \quad (1)$$

Extracted binary features are converted to decimal value and histogram of the constructed new image is considered

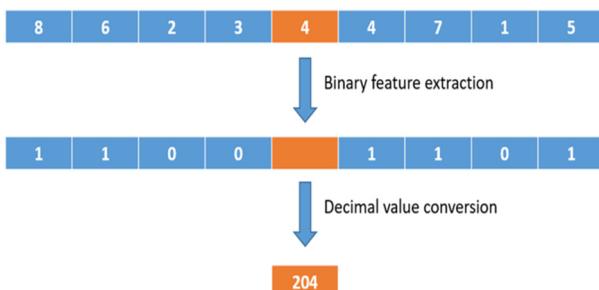


FIGURE 1. Graphical demonstration of the LBP. The histogram of the LBP applied image is utilized as feature vector.

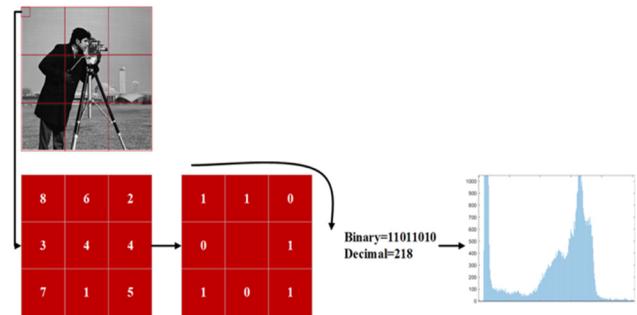


FIGURE 2. Graphical example of the one-dimensional local binary pattern.

as feature vector. It has also one dimensional version for the signal and voice processing [32].

As seen from Fig. 2, 9 sized overlapping blocks are used in 1D-LBP to extract features. In the binary feature extraction phase, 8 bits are extracted. Therefore, LBP extracts 256 features [16], [33].

B. TERNARY PATTERN

Local Ternary Pattern (LTP) is one of the LBP like feature extractor [34]. The main difference between LBP and LTP is to use ternary function for binary feature extraction [35]. Mathematical definition of the ternary pattern is given in Eq. 2 [18].

$$ter(IM_{t,k}, \text{center}, thr)$$

$$= tr = \begin{cases} 1, & IM_{t,k} - \text{center} > thr \\ 0, & -thr \leq IM_{t,k} - \text{center} \leq thr \\ -1, & IM_{t,k} - \text{center} < -thr \end{cases} \quad (2)$$

where tr represents ternary value and thr represents the threshold value.

Eq. 2 shows that the generated values are -1, 0 and 1 by using ternary function. To convert bits to ternary values, Eqs. 3 and 4 are used.

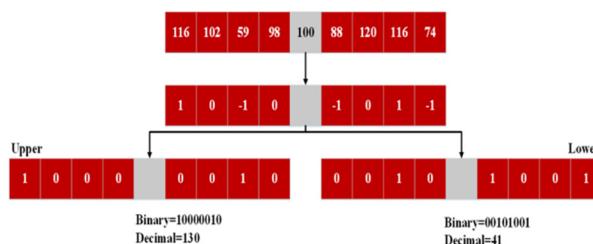
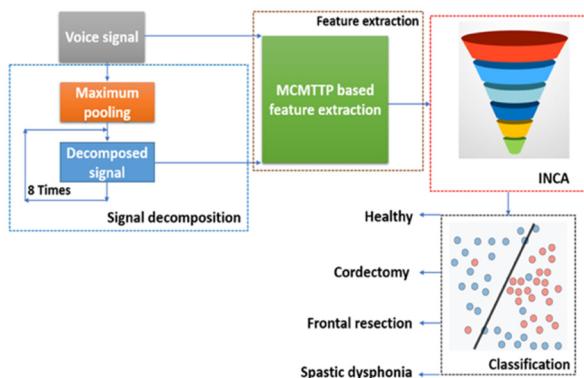
$$bit^{upper} = \begin{cases} 0, & tr < 1 \\ 1, & tr = 1 \end{cases} \quad (3)$$

$$bit^{lower} = \begin{cases} 0, & tr > -1 \\ 1, & tr = -1 \end{cases} \quad (4)$$

As seen from Eq. 3-4, the 8 lower and 8 upper bits are generated from the secondary layer (Eq.2). By converting these bits to decimal values, upper and lower signals are generated. Histograms of these signals are concatenate, and the 512 features are obtained by using TP. A graphical explanation of the TP is shown in Fig. 3 [36], [37].

III. THE PROPOSED MULTI CENTER AND MULTI THRESHOLD BASED TERNARY PATTERN: MCMTTP

As we have seen in section 2.1, LTP is efficient and easy to use for feature extraction for binary classification. Here we propose a novel feature extraction called MCMTTP which uses

**FIGURE 3.** A schematic demonstration of the TP.**FIGURE 4.** Graphical demonstration of the MCMTP and INCA based pathologic voice classification method.

simultaneously variable center value and variable threshold value to extract final feature from the image to improve the performance of LTP. The pseudo code of TP is described in Algorithm 1.

IV. THE PROPOSED VOICE-BASED DISEASE CLASSIFICATION METHOD

In this work, we propose an algorithm (MCMTP) whose goal is to discriminate between pathological and healthy voices with high accuracy rate by using multilevelled based feature extraction, hybrid feature selection and classification. The Our approach comprises three components: (1) Multilevelled feature extraction, (2) iterative NCA based feature selection and (3) a panoply of classifier for prediction. Flowchart of the algorithm is represented in Fig. 4 and detailed in the following section A, B and C.

A. MULTILEVELED FEATURE EXTRACTION USING MCMTP

In this phase, a novel multilevelled feature extraction network is proposed. To create features at different levels, maximum pooling method, down-sample technique to reduce features, is applied [38], [39]. This method has nine levels. In each level, center and threshold values are iteratively changed. Low, medium and high-level features are extracted by using the proposed multilevelled MCMTP feature extraction method. Mathematical explanations and steps of the proposed multilevelled method are given as below. Also, schematically explanation of the proposed MCMTP method is shown in Fig. 4.

Algorithm 1 Procedure of Parametric TP

```

Procedure: Parametric TP (PTP(signal, mt, center))
Input: Voice signal with size of M, mt is threshold value
multiplier parameter, center is center parameter. Output:
Feature vector (featvec) with length of 512.

00: thr = std (signal) * mt / 10; // Calculate threshold value
using standard deviation and multiplier parameter.
01: for i = 1 to M-8 do
02:   window = block(i : i + 8); // Divide voice signal
      into 9 sized overlapping blocks with size of 9.
03:   uppervalue (i) = 0; lowervalue (i) = 0 // Assign
      first values as 0.
04:   say = 1; // Define counter
05:   for j = 1 to 9 do
06:     if j != center then // Skip center value.
07:       tr = ter (block(j), center); // Calculate ternary
      values.
08:       Calculate upper and lower bits by using
      Eqs. 2-3.
09:       uppervalue (i) += bitupper (say) * 28-say;
      // Calculate upper values by using binary to decimal
      conversion.
10:       lowervalue (i) += bitlower (say) * 28-say;
      // Calculate lower values by using binary to decimal
      conversion.
11:       say += ; // Increase counter.
12:   end if
13:   end for j
14: end for i
15: Extract histograms of the uppervalue and lowervalue
signals.
16: Apply histogram concatenation to obtain featvec

```

Step 1: Create a loop from 1 to 9

Step 2: Extract features by using parametric TP function (see algorithm 1). TP is a textural feature extractor where it is used for images and signals as well. In the images, texture classification and analysis are one of the most studied areas of research. By using TP, 512 textural features are generated from a signal as follows.

$$\text{feature}^i = \text{PTP}(\text{signal}, i, 10 - i), \quad i = \{1, 2, \dots, 9\} \quad (5)$$

Step 3: Apply maximum pooling to signal by using Algorithm 2 [40].

$$\text{signal}^i = \text{maxpool}(\text{signal}) \quad (6)$$

Step 5: Repeat steps 2 and 3 nine times.

Step 6: Concatenated extracted features and obtain 4608 sized feature vector.

B. ITERATIVE NCA BASED FEATURE SELECTION

NCA is a distance based feature selection method which generates positive weight for each feature [41]. One of the

Algorithm 2 Procedure Maximum Pooling

Procedure: Maximum pooling ($\text{maxpool}(\text{signal})$)
Input: Voice signal with size of M
Output: New voice signal (signal^{new}) with length of M/2.

```

00: counter = 1; // Define counter to define new signal
index.
01: for i = 1 to M-1 step by 2 do
02:    $\text{signal}^{new}(\text{counter}) = \max(\text{signal}(i : i + 1))$ ; // Calculate maximum value of the divided non-overlapping
blocks.
03:   counter ++; // Increase counter.
04: end for i
```

limitations of NCA is that it can select redundant features as all features have positive weights and cannot detect optimal features [40]. To overcome this limitation, we propose an iterative NCA to be able to select optimal and non-redundant features. Therefore, a feature range is defined, (we choose in this paper the feature range between 40 to 1000). Similar to ReliefF [42], NCA weights are calculated, sorted by descending order. At each iteration, features using range values are proposed and a loss values is calculated (here we use KNN as a loss function). The features with the smallest error value are selected. This process defines our Iterative NCA,. Steps of our automated feature selection process INCA is presented as follows::

Step 1: Normalize data. We normalize 4608 sized feature vector using min-max normalization

$$X = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (7)$$

where X is the final feature vector with size of 4608. NCA is distance-based method and it uses stochastic gradient descent to update weights. Therefore, normalization should be applied to obtain optimal results.

Step 2: Generate NCA weights by using Eq. 8.

$$\text{weights} = \text{NCA}(X, \text{target}) \quad (8)$$

Step 3: Use NN to calculate error value. In this work, NN with Manhattan distance is used.

Step 4: Sort weights by descending.

$$[\text{sorted}_{\text{weights}} \text{endex}] = \text{sort}(\text{weights}) \quad (9)$$

Step 5: Select features using range value.

$$\begin{aligned}
\text{feat}^{k-39} &= \text{feat}(:, k) \\
&= X(:, \text{endex}(k)), k \\
&= \{1, 2, \dots, [40, 1000]\}
\end{aligned} \quad (10)$$

Step 6: Calculated error values of feat^n .

$$\text{error}(n) = \text{1NN}(\text{feat}^n, 10) \quad (11)$$

To calculate errors, 10-fold cross validation is used.

Step 7: Calculate minimum error.

$$[\text{minimum}, \text{index}] = \min(\text{error}) \quad (12)$$

Step 8: Select optimal feature (feat^{final}) by using calculated index value.

$$\begin{aligned}
\text{feat}^{final}(k) &= X(:, \text{endex}(k)), k \\
&= \{1, 2, \dots, \text{index} + 39\}
\end{aligned} \quad (13)$$

The generated feat^{final} is forwarded to classifiers.

C. CLASSIFICATION

In the classification phase, six widely conventional classifiers were used. These classifiers are naïve bayes (NB), k nearest neighborhood (kNN), linear discriminant (LD), decision tree (DT), support vector machine (SVM) and bagged tree (BT). 10-fold cross validation were selected as testing and training strategy. We used conventional classifiers to obtain a cognitive method and show success of the proposed MCMTTP and INCA based feature selector.

1) NB

NB is a widely used simple probability-based conventional classifier [43]. In this paper, for NB hyper parameter, we use Gaussian kernel [44].

2) KNN

KNN is a distance based parametric classifier and is one of the simplest classifiers in the literature. In this paper, k and distance metrics are selected as 1 and Euclidean respectively.

3) LD

LD is a basic linear discriminant classifier based on the mean and the covariance of each class to where we need to classify a data point. Therefore, any hyper parameter setting is not used [45].

4) DT

DT or decision tree classifier and is one of the widely preferred conventional classifiers and uses information entropy to classify observations. Some of the popular tree algorithms are C4.5, CART and ID3. The widely used split criterion of the DT are Gini, Twoing and maximum deviance reduction. It is a parametric classifier. In this work, Gini's model is used in DT [46].

5) SVM

SVM or support vector machine is an optimization based conventional classifier that incorporates a variety of kernel methods such as radial basis sets, polynomial kernel or neural networks. We selected 3rd degree polynomial kernel or what is called cubic SVM [47].

6) BT

BT or bagged tree is an ensemble model of the DT. The used hyper parameters in BT are given as follows. Ensemble

method is bag and Gini's diversity index is used for random split selection [48].

The conspicuous attributes of the proposed method are summarized as follows:

- Generating a comprehensively features by using MTM-CTTP method.
- A multileveled feature generating network is proposed to extract low, medium and high-level features by using MTMCTTP method and maximum pooling technique.

To denote a general success and strength of the proposed feature generation network and INCA feature selector, six traditional classifiers are used.

V. PERFORMANCE ANALYSIS

To evaluate performance of the proposed MCMCTTP based voice classification method, Saarbruecken Voice Database (SVD) [14], [49] was used. This dataset has pathological voices and user can freely download these voices. These pathologic voices were collected from more than 2000 subjects in 70 classes (diseases) with 50 KHz frequency and 16-bit resolution. These voices were collected as vowels (/a/, /i/, /u/) and sentence ("Guten Morgen, ne es Ihnen?"). Three diseases are used to create cases and these diseases are frontal resection, cordeectomy and spastic dysphonia. We collected sentences from SVD. Eight cases were defined. The defined cases are given as below.

Case 1: In this case, two classes are used. These classes are healthy and cordeectomy. The proposed INCA selected 24 features to solve this classification problem with high accuracy.

Case 2: In this case, two classes are used. These classes are healthy and frontal resection. The proposed INCA selected 267 features to solve this classification problem with high accuracy.

Case 3: In this case, two classes are used. These classes are healthy and spastic dysphonia. The proposed INCA selected 404 features to solve this classification problem with high accuracy.

Case 4: In this case, two classes are used. These classes are named as healthy and patient. There are voices of cordeectomy, frontal resection and spastic dysphonia voices in the patient class. The proposed INCA selected 57 features to solve this classification problem with high accuracy.

Case 5: In this case, three classes are used. These classes are healthy, cordeectomy and frontal resection. The proposed INCA selected 144 features to solve this classification problem with high accuracy.

Case 6: In this case, three classes are used. These classes are healthy, cordeectomy and spastic dysphonia. The proposed INCA selected 36 features to solve this classification problem with high accuracy.

Case 7: In this case, three classes are used. These classes are healthy, frontal resection and spastic dysphonia. The proposed INCA selected 384 features to solve this classification problem with high accuracy.

Case 8: In this case, four classes, healthy, cordeectomy, frontal resection and spastic dysphonia, are considered. A high accuracy has been achieved when 71 features were selected by the proposed INCA.

To study the performance of those classifiers, classification accuracy and geometric mean are used. These metrics are mathematically shown in below [50].

$$\text{accuracy} = \frac{tp + tn}{tp + tn + fp + fn} \quad (14)$$

$$gm = \sqrt{\frac{tp}{tp + fn} \times \frac{tn}{tn + fp}} \quad (15)$$

where tp , tn , fp and fn are true positives, true negatives, false positives and false negatives. The calculated results are listed in Tables 2 and 3.

TABLE 2. Accuracy values (%) of the proposed MCMCTTP and INCA voice-based disease detection method according to cases a.

	kNN	SVM	DT	LD	NB	BT
Case 1	99.03	99.03	97.41	97.73	99.35	98.38
Case 2	100.0	100.0	98.60	92.28	97.54	98.60
Case 3	97.13	96.82	92.28	90.13	88.85	94.27
Case 4	98.28	97.30	93.87	92.40	93.63	94.85
Case 5	91.86	93.31	87.79	96.63	87.79	89.24
Case 6	94.64	94.64	87.67	92.49	90.35	92.23
Case 7	95.70	96.56	87.97	79.08	89.11	91.69
Case 8	87.99	89.95	81.37	84.80	81.37	84.07

TABLE 3. Geometric mean values (%) of the proposed MCMCTTP and INCA based voice-based disease detection method according to cases a.

	kNN	SVM	DT	LD	NB	BT
Case 1	97.42	98.09	94.40	94.59	98.95	97.03
Case 2	100.0	100.0	95.43	83.92	97.37	95.43
Case 3	95.24	93.04	87.64	86.65	87.14	88.37
Case 4	98.13	96.97	93.08	90.94	93.02	94.25
Case 5	78.26	81.75	65.50	67.91	71.09	67.54
Case 6	91.13	91.09	81.08	87.48	85.61	86.44
Case 7	94.0	93.35	81.38	70.01	86.51	83.19
Case 8	76.97	80.93	65.39	71.26	73.05	73.38

As seen from tables 2 and 3, Case 2 achieved 100.0% classification accuracy and geometric mean by using kNN and SVM. There are two classes in the cases 1-4. Their best classification accuracies were at least 97%. The calculated best geometric mean achieved more than 95% for cases 1-4. The chosen classifier we obtained is kNN for all cases. Cases 5-7 have three classes. Our proposed provided 90% or more classification rates for these cases. However, the best obtained geometric mean for Case 5 was obtained and is 81.75%.

As we have seen, we used heterogeneous dataset to study the performance of our algorithm. For Case 8, 4 classes are used and the voice-based diseases detection with kNN and

TABLE 4. Geometric mean (%) values of the proposed method and other methods.

Voice	Method	Case	Geometric mean
/a/ vowel normal	Zhang et al.'s first method [51]	Case 1	81.46
/a/ vowel normal	Tuncer et al.'s method [16]	Case 1	86.23
/a/ vowel normal + low high low	Zhang et al.'s second method [51]	Case 1	75.30
/a/ vowel normal + low high low	Concatenate based method [17]	Case 1	80.14
/a/ vowel normal + low high low	Wu et al.'s method [17]	Case 1	85.86
Sentence	Our method + SVM	Case 1	98.09
/a/ vowel normal	Zhang et al.'s first method [51]	Case 2	93.49
/a/ vowel normal	Tuncer et al.'s method [16]	Case 2	85.83
/a/ vowel normal + low high low	Zhang et al.'s second method [51]	Case 2	93.25
/a/ vowel normal + low high low	Concatenate based method [17]	Case 2	91.06
/a/ vowel normal + low high low	Wu et al.'s method [17]	Case 2	93.44
Sentence	Our method + SVM	Case 2	100.0

other classifiers was challenging except when using SVM which provided A 89.95% accuracy and 80.93% geometric mean.

VI. DISCUSSIONS

In this study, we proposed a novel multileveled feature extraction method which is a modified version of the TP called MCMTTP. To create MCMTTP, a parametric TP function is presented. Nine feature extraction networks are created using maximum pooling to extract several features at different levels. On the other hand, we use an effective feature selector INCA to select optimal features. We investigated the performance of the proposed algorithm on several pathologic voice dataset (SVD). Eight several cases were created by using these voices. In the literature, machine learning based voice disease detection methods have been applied to discriminate two class. Here we proposed an algorithm that classify 3 and 4 classes using also heterogeneous dataset. Tables 1 and 2 summarizes the performance of our algorithm and shows that the best classifier are kNN and SVM with more a classification rate greater than 90% of all cases except for Case 8 in which 80.93% geometric mean was obtained. To show the success of the proposed MCMTTP based method, results of the other methods were listed in Table 4.

As seen from classification results, the proposed MCMTTP based voice classification method achieved higher geometric mean than other selected works. Table 4 compares our algorithm with other approaches for case 1 and 2 only

(binary). We could not do a comparative study between our algorithm and other classes for nonbinary. We summarize the advantages of the proposed MCMTTP, and INCA based method as the followings:

- 1) Automated feature selection process is solved by using INCA.
- 2) By using a modified TP (MCMTTP) and maximum pooling method, a multilevelled feature extraction method is proposed, and high accuracy rates were achieved by using this feature extraction and INCA based feature selection method together.
- 3) Six conventional classifiers were used to show strength of the proposed voice-based disease detection method.
- 4) Eight cases were defined to obtain general results.
- 5) The proposed MCMTTP and INCA based pathologic voice classification method outperforms.

VII. CONCLUSION

In this work, a novel one-dimensional improved TP method was presented along with an automated voice-based disease classification method. In this method, a multilevelled feature extraction network is constructed where features are extracted from each level by parametric TP. The extracted features are concatenated, and the proposed INCA selects discriminative ones. We used SVD pathologic voice dataset to create 8 cases. We selected three diseases for these cases. These are corpectomy, frontal resection and spastic dysphonia. This method reached 100.0% classification accuracy and geometric mean (perfect classification) for Case 2 (frontal resection detection). The proposed MCMTTP and INCA based method achieved high performance (See tables 2-3). This method also achieved better results than other selected presented methods (See Table 4). These results clearly demonstrated that the proposed MCMTTP and INCA method is successful. The proposed MCMTTP and INCA methods are cognitive and lightweight method. Therefore, our techniques are useful for most of new generation of real time applications like a chest diseases clinic as smart assistants.

ACKNOWLEDGMENT

The authors would like to thank Qatar National Library, QNL, for supporting us in publishing our research.

REFERENCES

- [1] P. Gómez-Vilda, V. Rodellar-Biarge, V. Nieto-Lluis, C. Muñoz-Mulas, L. M. Mazaira-Fernández, C. Ramírez-Calvo, M. Fernández-Fernández, and E. Toribio-Díaz "Neurological disease detection and monitoring from voice production," in *Proc. Int. Conf. Nonlinear Speech Process.*, 2011, pp. 1-8.
- [2] L. Naranjo, C. J. Pérez, Y. Campos-Roca, and J. Martín, "Addressing voice recording replications for Parkinson's disease detection," *Expert Syst. Appl.*, vol. 46, pp. 286–292, Mar. 2016.
- [3] N. A. Saloni, R. K. Sharma, and A. K. Gupta, "Disease detection using voice analysis: A review," *Int. J. Med. Eng. Informat.*, vol. 6, no. 3, p. 189, 2014.
- [4] L. Naranjo, C. J. Pérez, J. Martín, and Y. Campos-Roca, "A two-stage variable selection and classification approach for Parkinson's disease detection by using voice recording replications," *Comput. Methods Programs Biomed.*, vol. 142, pp. 147–156, Apr. 2017.

- [5] M. B. Hoy, "Alexa, siri, cortana, and more: An introduction to voice assistants," *Med. Reference Services Quart.*, vol. 37, no. 1, pp. 81–88, Jan. 2018.
- [6] G. López, L. Quesada, and L. A. Guerrero, "Alexa vs. Siri vs. Cortana vs. Google Assistant: A comparison of speech-based natural user interfaces," in *Proc. Int. Conf. Appl. Hum. Factors Ergonom.*, 2017, pp. 241–250.
- [7] R. A. Franco and J. G. Andrus, "Common diagnoses and treatments in professional voice users," *Otolaryngologic Clinics North Amer.*, vol. 40, no. 5, pp. 1025–1061, Oct. 2007.
- [8] E. D. Übeyli and İ. Güler, "Feature extraction from Doppler ultrasound signals for automated diagnostic systems," *Comput. Biol. Med.*, vol. 35, no. 9, pp. 735–764, Nov. 2005.
- [9] R. A. Miller, "Medical diagnostic decision support systems-past, present, and future: A threaded bibliography and brief commentary," *J. Amer. Med. Inform. Assoc.*, vol. 1, no. 1, pp. 8–27, Jan. 1994.
- [10] K. Hajian-Tilaki, "Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation," *Caspian J. Internal Med.*, vol. 4, no. 2, p. 627, 2013.
- [11] E. D. Übeyli, "Implementing automated diagnostic systems for breast cancer detection," *Expert Syst. Appl.*, vol. 33, no. 4, pp. 1054–1062, Nov. 2007.
- [12] S. Fujimura, T. Kojima, Y. Okanoue, K. Shoji, M. Inoue, K. Omori, and R. Hori, "Classification of voice disorders using a one-dimensional convolutional neural network," *J. Voice*, Mar. 2020.
- [13] I. Hammami, L. Salhi, and S. Labidi, "Voice pathologies classification and detection using EMD-DWT analysis based on higher order statistic features," *IRBM*, Jan. 2020.
- [14] W. Barry and M. Pützer, "Saarbrücken voice database," Inst. Phonetics, Tech. Rep., 2007.
- [15] J. M. Tracy, Y. Özkanca, D. C. Atkins, and R. Hosseini Ghomi, "Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson's disease," *J. Biomed. Informat.*, vol. 104, Apr. 2020, Art. no. 103362.
- [16] T. Tuncer, S. Dogan, and F. Ertam, "Automatic voice based disease detection method using one dimensional local binary pattern feature extraction network," *Appl. Acoust.*, vol. 155, pp. 500–506, Dec. 2019.
- [17] K. Wu, D. Zhang, G. Lu, and Z. Guo, "Joint learning for voice based disease detection," *Pattern Recognit.*, vol. 87, pp. 130–139, Mar. 2019.
- [18] T. Tuncer and S. Dogan, "Novel dynamic center based binary and ternary pattern network using m4 pooling for real world voice recognition," *Appl. Acoust.*, vol. 156, pp. 176–185, Dec. 2019.
- [19] T. T. Miller, "MR imaging of the knee," *Sports Med. Arthroscopy Rev.*, vol. 17, no. 1, pp. 56–67, 2009.
- [20] D. Pravena and S. Dhivya, "Pathological voice recognition for vocal fold disease," *Int. J. Comput. Appl.*, vol. 47, no. 13, pp. 31–37, 2012.
- [21] A. U. Haq, J. P. Li, M. H. Memon, J. Khan, A. Malik, T. Ahmad, A. Ali, S. Nazir, I. Ahad, and M. Shahid, "Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings," *IEEE Access*, vol. 7, pp. 37718–37734, 2019.
- [22] T. Bocklet, E. Noth, G. Stemmer, H. Ruzickova, and J. Rusz, "Detection of persons with Parkinson's disease by acoustic, vocal, and prosodic analysis," in *Proc. IEEE Workshop Autom. Speech Recognit. Understand.*, Dec. 2011, pp. 478–483.
- [23] J. Rusz, R. Cmejla, H. Ruzickova, and E. Ruzicka, "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease," *J. Acoust. Soc. Amer.*, vol. 129, no. 1, pp. 350–367, 2011.
- [24] W. Caesarendra, M. Ariyanto, J. D. Setiawan, M. Arozi, and C. R. Chang, "A pattern recognition method for stage classification of Parkinson's disease utilizing voice features," in *Proc. IEEE Conf. Biomed. Eng. Sci. (IECBES)*, Dec. 2014, pp. 87–92.
- [25] Parkinson's Telemonitoring Data Set by the University of CaliforniaIrvine University (UCI). [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring>
- [26] A. Benba, A. Jilbab, and A. Hammouch, "Using human factor cepstral coefficient on multiple types of voice recordings for detecting patients with Parkinson's disease," *IRBM*, vol. 38, no. 6, pp. 346–351, Nov. 2017.
- [27] M. Hariharan, K. Polat, and R. Sindhu, "A new hybrid intelligent system for accurate detection of Parkinson's disease," *Comput. Methods Programs Biomed.*, vol. 113, no. 3, pp. 904–913, Mar. 2014.
- [28] T. Ojala, M. Pietikäinen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in *Proc. 12th Int. Conf. Pattern Recognit.*, 1994, pp. 582–585.
- [29] T. Ojala, M. Pietikäinen, and T. Mäenpää, "A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification," in *Proc. Int. Conf. Adv. Pattern Recognit.*, 2001, pp. 399–408.
- [30] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657–1663, Jun. 2010.
- [31] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 469–481.
- [32] Y. Kaya, M. Uyar, R. Tekin, and S. Yıldırım, "1D-local binary pattern based feature extraction for classification of epileptic EEG signals," *Appl. Math. Comput.*, vol. 243, pp. 209–219, Sep. 2014.
- [33] L. Houam, A. Hafiane, A. Boukrouche, E. Lespessailles, and R. Jennane, "One dimensional local binary pattern for bone texture characterization," *Pattern Anal. Appl.*, vol. 17, no. 1, pp. 179–193, Feb. 2014.
- [34] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *Proc. Int. Workshop Anal. Modeling Faces Gestures*, 2007, pp. 168–182.
- [35] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.
- [36] T. Tuncer, S. Dogan, and A. Subasi, "Surface EMG signal classification using ternary pattern and discrete wavelet transform based feature extraction for hand movement recognition," *Biomed. Signal Process. Control*, vol. 58, Apr. 2020, Art. no. 101872.
- [37] L. Nanni, S. Braham, and A. Lumini, "A local approach based on a local binary patterns variant texture descriptor for classifying pain states," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 7888–7894, Dec. 2010.
- [38] D. Yu, H. Wang, P. Chen, and Z. Wei, "Mixed pooling for convolutional neural networks," in *Proc. Int. Conf. Rough Sets Knowl. Technol.*, 2014, pp. 364–375.
- [39] D. Miao, W. Pedrycz, D. Slezak, G. Peters, Q. Hu, and R. Wang, "Rough sets and knowledge technology," in *Proc. 9th Int. Conf. RSKT*, vol. 8818. Shanghai, China: Springer, Oct. 2014.
- [40] Y.-D. Zhang, K. Muhammad, and C. Tang, "Twelve-layer deep convolutional neural network with stochastic pooling for tea category classification on GPU platform," *Multimedia Tools Appl.*, vol. 77, no. 17, pp. 22821–22839, Sep. 2018.
- [41] S. Raghu and N. Sriraam, "Classification of focal and non-focal EEG signals using neighborhood component analysis and machine learning algorithms," *Expert Syst. Appl.*, vol. 113, pp. 18–32, Dec. 2018.
- [42] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, nos. 1–2, pp. 23–69, Oct. 2003.
- [43] J. Chen, H. Huang, S. Tian, and Y. Qu, "Feature selection for text classification with Naïve bayes," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5432–5435, Apr. 2009.
- [44] I. Rish, "An empirical study of the naive Bayes classifier," in *Proc. IJCAI Workshop Empirical Methods Artif. Intell.*, 2001, pp. 41–46.
- [45] S. Chen and X. Yang, "Alternative linear discriminant classifier," *Pattern Recognit.*, vol. 37, no. 7, pp. 1545–1547, Jul. 2004.
- [46] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 3, pp. 660–674, May/Jun. 1991.
- [47] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, "A fast iterative nearest point algorithm for support vector machine classifier design," *IEEE Trans. Neural Netw.*, vol. 11, no. 1, pp. 124–136, Jan. 2000.
- [48] T. Hothorn and B. Lausen, "Bundling classifiers by bagging trees," *Comput. Statist. Data Anal.*, vol. 49, no. 4, pp. 1068–1078, Jun. 2005.
- [49] D. Martínez, E. Lleida, A. Ortega, A. Miguel, and J. Villalba, "Voice pathology detection on the Saarbrücken voice database with calibration and fusion of scores using MultiFocal toolkit," in *Advances in Speech and Language Technologies for Iberian Languages*. Cham, Switzerland: Springer, 2012, pp. 99–109.

- [50] T. Tuncer, S. Dogan, P. Plawiak, and U. R. Acharya, "Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals," *Knowl.-Based Syst.*, vol. 186, Dec. 2019, Art. no. 104923.
- [51] Z. Zhang, Z. Lai, Y. Xu, L. Shao, J. Wu, and G.-S. Xie, "Discriminative elastic-net regularized linear regression," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1466–1481, Mar. 2017.



TURKER TUNCER was born in Tunceli, Turkey, in 1986. He received the B.S. degree from the Department of Electronics and Computer Education, Technical Education Faculty, Firat University, in 2009, the M.S. degree in telecommunication science from Firat University, in 2011, and the Ph.D. degree from the Department of Software Engineering, Firat University, in 2016. He is currently an Assistant Professor with the Department of Digital Forensics Engineering, Firat University. He has more than 80 articles in national, international journals and conferences. His research interests include feature engineering, image processing, signal processing, data hiding, image authentication, cryptanalysis, and cryptography.



SENGUL DOGAN received the Ph.D. degree in electrical and electronic engineering from the University of Firat, Elazığ, Turkey, in 2011. She is currently an Associate Professor with the Digital Forensics Engineering Department, Firat University. Her research interests cover data hiding, information security, digital forensics, image processing, signal classification, and optimization techniques.



FATİH ÖZYURT was born in Erzurum, Turkey. He received the bachelor's degree in computer engineering from Eastern Mediterranean University, Cyprus, in 2011, the master's degree in computer engineering from Fatih University, Istanbul, in 2014, and the Ph.D. degree in software engineering from Firat University, Elazığ, in 2019. He is currently an Assistant Professor with the Department of Software Engineering, Firat University. His research interests include pattern recognition, artificial neural networks, image processing, deep learning, and information security.



SAMIR BRAHIM BELHAOUARI received the master's degree in telecommunications and network from the Institute National Polytechnique of Toulouse, France, in 2000, and the Ph.D. degree in mathematics from the Federal Polytechnic School of Lausanne-Switzerland, in 2006. He is currently an Associate Professor with the University of Hamad Bin Khalifa, Qatar foundation, in the Division of Information and Communication Technologies, College of Science and Engineering, HBKU. During last years, he also holds several Research and Teaching positions at Innopolis University-Russia, Alfaisal University-Saudi Arabia, the University of Sharjah-UAE, University Technology PETRONAS-Malaysia, and the EPFL Federal Swiss School-Switzerland. His research interests include vary areas from applied mathematics, statistics, and data analysis until artificial intelligence, image and signal processing (biomedical, bioinformatics, forecasting, etc.), due to both Mathematics and Computer Science backgrounds.



HALIMA BENSMAIL received the Ph.D. degree from the University of Pierre and Marie Curie (Paris 6), France, spend one year as a Postdoctoral Researcher at the University of Washington (UW), Seattle, and one year at the Fred Hutchinson Cancer Research Center (FHCRC), held two years as a Research Scientist at the Data Theory Group, University of Leiden, The Netherlands. She held an Assistant and Associate Professor of statistical machine learning at the University of Tennessee in statistics, in 2001, and an Associate Professor of biostatistics at the University of Virginia, in 2006. She is currently a Principal Scientist with the Qatar Computing Research Institute of Bioinformatics and Computational Biology and a Joint Associate Professor with the College of Computer and Science Engineering, HBKU. She has published more than 80 peer-reviewed articles in high impact journal, such as *Nucleic Acids Research*, *Genome Research*, *Bioinformatics*, *Briefings in Bioinformatics*, *Pattern Recognition*, and the *Journal of American Statistical Association*. She received several distinguished awards, such as the International Federation of Classification in Society Award, in 1997, and the Qatar Foundation Computing Research Award, in 2011.

• • •