Pertanika J. Sci. & Technol. 28 (3): 839 - 858 (2020)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Dual-Tree Complex Wavelet Packet Transform for Voice Pathology Analysis

Farah Nazlia Che Kassim[1]*, Hariharan Muthusamy[2], Vikneswaran Vijean[1], Zulkapli Abdullah[3] and Rokiah Abdullah[1]

[1]School of Mechatronic Engineering, Universiti Malaysia Perlis, Kampus Pauh Putra, 02600 UniMAP, Arau, Perlis, Malaysia
[2]Department of Biomedical-Engineering, SRM Institute of Science and Technology, SRM Nagar, Kattankulathur 603203 Kancheepuram District, Tamil Nadu, India
[3]Pusat Kejuruteraan, Universiti Malaysia Perlis, Kampus Pauh Putra 02600 UniMAP, Arau, Perlis, Malaysia

ABSTRACT

Voice pathology analysis has been one of the useful tools in the diagnosis of the pathological voice, as the method is non-invasive, inexpensive, and can reduce the time required for the analysis. This paper investigates feature extraction based on the Dual-Tree Complex Wavelet Packet Transform (DT-CWPT) using energy and entropy measures tested with two classifiers, k-Nearest Neighbors (k-NN) and Support Vector Machine (SVM). Massachusetts Eye and Ear Infirmary (MEEI) voice disorders database and Saarbruecken Voice Database (SVD) were used. Five datasets of voice samples were used from these databases, including normal and abnormal samples, Cysts, Vocal Nodules, Polyp, and Paralysis vocal fold. To the best of the authors' knowledge, very few studies were done on multiclass classifications using specific pathology database. File-based and frame-based investigation for two-class and multiclass were considered. In the two-class analysis using the DT-CWPT with entropies, the classification accuracy of 100% and 99.94% was achieved for MEEI and SVD database respectively. Meanwhile, the classification accuracy for multiclass analysis comprised of 99.48% for the MEEI database and 99.65% for SVD database. The experimental results using the proposed features provided promising accuracy to detect the presence of diseases in vocal fold.

Keywords: Dual-tree complex wavelet packet transform, file-based, frame-based, two-class and multiclass, voice pathology analysis

ARTICLE INFO

Article history:
Received: 24 February 2020
Accepted: 18 March 2020
Published: 16 July 2020

E-mail addresses:
nazlia@unimap.edu.my (Farah Nazlia Che Kassim)
hariharan.m@ktr.srmuniv.ac.in (Hariharan Muthusamy)
vikneswaran@unimap.edu.my (Vikneswaran Vijean)
zulkapli@unimap.edu.my (Zulkapli Abdullah)
rokiah@unimap.edu.my (Rokiah Abdullah)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

Pathological changes of the larynx are presented by the failure of the vocal fold to move continuously and properly, which can affect the voice. Voice changes may include loss of power, changes in the pitch, constriction of the voice range (i.e. displacement towards lower frequency), the addition of noises, and others (Vikram & Umarani, 2013). The precise laryngeal diagnostic methods like endoscopy and laryngoscopy used in clinical practice can cause discomfort to the patient, invasive, and expensive. By this reason, detection of the disease in its early stage is required. A precise voice signal diagnostic quantitative and non-invasive nature allows the identification and monitoring of vocal fold pathology, as well as reducing the time and cost required for detection.

Patient voice recording allows researchers to analyse a variety of parameters. The acoustics features identify the pathology based on the functioning and condition of various speech organs such as fundamental frequency, jitter, shimmer, harmonic to noise ratio, and intensity (Teixeira et al., 2013). A long duration of the signal is needed to extract the features in the time domain, which is tough to get from affected patients. For this reason, researchers start to explore the frequency domain analysis, which requires less data that offers more information. Mel Frequency Cepstral Coefficients (MFCC) has been reported as a very successful parameter for pathological voice detection (Srinivasan et al., 2014). Although MFCC is renowned and widely used, some limitations exist, such as low robustness to noisy signals (Harar et al., 2018). Mekyska et al. (2015), who studied the parameterisation techniques based on segmental features, such as MFCC and Linear Predictive Coding (LPC), provided the best classification results of 82.1%–100%. These techniques were, however, usually challenging to be clarified clinically. Among the limitations raised were identification of particular voice diseases and detection in its first stage or evaluation of its progress.

In recent times, enormous interest has emerged in wavelets approaches for pathological voice detection. Wavelet Packet Transform (WPT) was found to be an excellent tool for the analysis of non-stationary signals both in time and frequency scale (Hariharan et al., 2014). Decomposing a signal into wavelets rather than frequencies can give a much better resolution in the domain it is transformed into. Although a great deal of literature exists concerning voice pathology analysis, only a handful of them had employed time-frequency analysis for the investigation of pathology detections. This study focuses on investigating the use of DT-CWPT for voice pathology analysis. DT-CWPT produces complex coefficients using a dual-tree of wavelet filters to obtain real and imaginary coefficients. This would be useful as an effort to identify new features that can contribute to the overall best performance to detect the specific pathology.

Related Works

Voice pathology analysis focuses on employing signal processing techniques and machine learning algorithms to form a system capable of precise and accurate detection. Wavelet decomposition for feature extraction has been one of the great approaches in this field. Most works were done using file-based analysis, where the whole audio file is considered as the input signal to further classified as normal or pathological. Very few studies have been carried out as frame-based (Godino-llorente et al., 2005; Hariharan et al., 2014). Hariharan et al. (2014) proposed a new feature vector based on the WPT and singular value decomposition using four differently supervised classifiers, such as k-NN, least-square SVM, probabilistic neural network, and general regression neural network. In their paper, 100% classification accuracy was attained using the proposed features and classifiers for normal and abnormal vocal fold detection in both MEEI and MAPACI speech pathology database. Akbari and Arjmandi (2014) explored the possibility of applying the Discrete Wavelet Packet Transform (DWPT) to categorise 258 voiced samples, randomly selected from three pathologic classes and one normal class in MEEI database. The disordered voice samples comprised hyperfunction, gastric reflux, and A–P squeezing. Feature vectors optimised using Multiclass Linear Discriminant Analysis showed an average performance of 96.67% and 97.33% for Energy and entropy features, respectively classified by Multilayer Neural Network. Saidi and Almasganj (2015) obtained a good classification rate of 99.3% for normal and abnormal cases classified by SVM using extracted features from a five-band wavelet system. Features were extracted from a total of 57 normal and 653 pathological voice signals in the MEEI database containing sustained vowel /ah/ and speech sample pronounced the "Rainbow passage". Majidnezhad (2015) explored an initial feature vector based on the combination of the Wavelet Packet Decomposition (WPD) and the MFCC using a hybrid of the Artificial Neural Network (ANN) as the classifier, which gave 94.24% accuracy on the MEEI database and 95.3% accuracy on the Russian database (RusDS).

While most of the current work focuses on distinguishing normal (healthy) and abnormal (pathological) voices using various parameters, very little research on the multiclass classification system of different types of pathologies have been conducted. A study by Muhammad et al. (2017) on multiclass experimental results indicated that the Interlaced Derivative Pattern (IDP) based features using SVM gave greater accuracy than those using conventional MFCC and Multi-Dimensional Voice Program parameters in three different databases, which are the MEEI, SVD, and Arabic Voice Pathology Database (AVPD). The proposed IDP based features using SVM achieved 99.38% (MEEI), 93.2% (SVD), and 91.5% (AVPD) average accuracies for two-class classification. However, it is a

challenging task to compare between published papers, since their findings varied because of the differences in the chosen voice pathology samples from different databases, acoustic features implemented, and classifiers employed in the researches.

DT-CWPT introduced by Bayram and Selesnick (2008) had several properties such as the introduction of limited redundancy, reduced aliasing, and nearly shift-invariance, which were lacking in conventional WPT. Recently, the successful application of DT-CWPT in various fields such as mechanical fault diagnosis (Qu et al., 2016; Cao et al., 2019; Haidong et al., 2019), infant cry classification (Lim et al., 2018), speaker, and accent recognition (Abdullah et al., 2019) has been reported. Since the wavelet packet analysis has a reliable capability in identifying vocal fold pathology, this study aimed to investigate the use of DT-CWPT for analysing the voice signals using energy and entropy measures. Two-class and multiclass experiments were performed using the file-based and frame-based approach. Five datasets of voice samples for two-class and multiclass analysis from two databases were used for the investigation so that a direct comparison could be made with that of the previous studies.

METHODOLOGY

Figure 1 shows the block diagram of the proposed voice pathology analysis in this study.

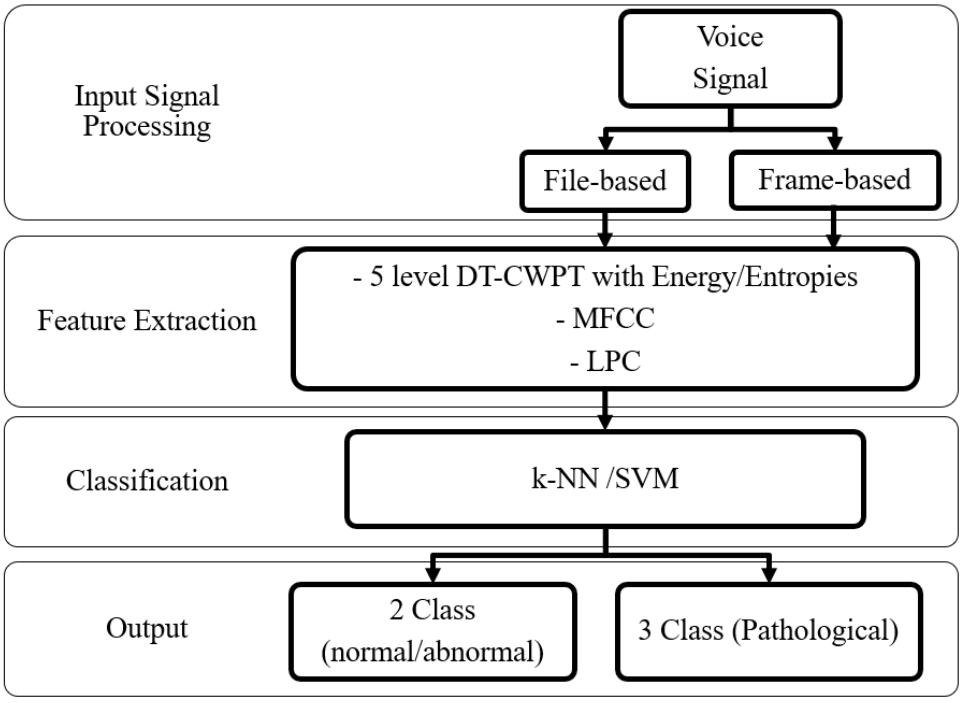

Figure 1. Block diagram of the proposed voice pathology analysis

Input Signal Processing

The voice signals from the normal person and patients suffering from disorders were acquired from the MEEI and SVD databases. Five datasets of voice samples were used from these databases, including normal and abnormal samples, Cysts, Vocal Nodules, Polyp, and Paralysis vocal folds. Table 1 shows the number of voice samples used as datasets for two-class and multiclass investigations.

Table 1

Number of samples for two-class and multiclass analysis

Dataset	Database	Analysis		Voice Sample		Total Samples
1	MEEI	Two-class	Class 1	Abnormal	173	226
			Class 2	Normal	53	
2	MEEI	Two-class	Class 1	Abnormal	106	159
			Class 2	Normal	53	
3	SVD	Two-class	Class 1	Abnormal	244	931
			Class 2	Normal	687	
4	MEEI	Multiclass	Class 1	Vocal nodules	19	106
			Class 2	Paralysis	67	
			Class 3	Polyp	20	
5	SVD	Multiclass	Class 1	Cysts	6	244
			Class 2	Paralysis	194	
			Class 3	Polyp	44	

The MEEI database, which is the most widely used and the only commercially available database, become a benchmark in the field of pathological speech analysis (Harar et al., 2018). Meanwhile, the SVD database, a freely downloadable database, was recorded by the Institute of Phonetics of Saarland University (Barry & Pützer, 2007). Only a few studies of voice pathology analysis have been explored in this database (Martinez et al., 2012; Muhammad et al., 2017). The voice signal files that only contained sustained normal pitch vowel /a/ samples were selected. All voice samples were down-sampled to have the same sampling frequency of 25 kHz due to the different recording sampling rates stored in this database. This rate was exploited because it satisfied the minimum rate specified by Nyquist and also the rate was mostly used in other referenced papers (Hariharan et al., 2014; Muhammad et al., 2017; Harar et al., 2018; Patil, 2019). This sample analysis was chosen to present outcomes comparable with previously published works. In the MEEI database, only samples of vocal nodules were available, while in SVD, cysts samples were

provided. Therefore, to investigate the three pathologies and to allow easier comparison between the MEEI and SVD databases, class 1 was represented to be either vocal nodules or cysts depending on the database employed.

However, unlike the previous works, file-based and frame-based analyses were conducted to produce a larger dataset. In the frame-based analysis, voice samples were segmented into frames of 40 ms long (as the voice are considered stationary in the period of 20–40 ms) using a Hamming window with 50% overlap (Shafik et al., 2009; Hariharan et al., 2014).

Feature Extraction

The DT-CWPT using energy and entropy measures are proposed as the feature extraction. It is an extended algorithm from the Dual-Tree Complex Wavelet Transform (DT-CWT), with two bands of DWPT operating in parallel (Bayram & Selesnick, 2008). The DT-CWT is a form of the discrete wavelet transform, which generates complex coefficients (real and imaginary) using a dual-tree of wavelet filters that offer a more productive signal analysis. This introduces limited redundancy (2m:1 for m-dimensional signals) and allows the transform to provide approximate shift-invariance and directionally selective filters (properties lacking in the traditional wavelet transform) while preserving the natural properties of perfect reconstruction and computational efficiency with good, well-balanced frequency responses (Selesnick et al., 2005). DT-CWPT has the same properties of DT-CWT i.e shift-invariance and excellent directional selectivity, with the advantage of fewer energy leakages into its negative frequency bands (Serbes et al., 2013).

As shown in Figure 2, each of the sub-bands should be repeatedly decomposed using low-pass/high-pass perfect reconstruction (PR) filter banks (FB) to construct DT-CWPT. The PR FBs should be chosen so that the response of each branch of the second wavelet packet FB is the discrete Hilbert transform of the corresponding branch of the first wavelet packet FB; thus, allowing each sub-band of the DT-CWPT to be analytic. The PR FB, which is used to decompose the first FB of the DT-CWT, should also be used to decompose the second FB to preserve the Hilbert transform relationship already satisfied by those branches. The high-pass branch of the first stage, $h_1^{(1)}(n)$ and $h'_1{}^{(1)}(n)$, satisfy $h'_1{}^{(1)}(n) = h_1^{(1)}(n-1)$, which is exactly the same relationship satisfied by the low-pass filters of the first stage, $h'_0{}^{(1)}(n) = h_0^{(1)}(n-1)$. The second wavelet packet FB is obtained by replacing the first stage filters $h_i^{(1)}(n)$ by $h_i^{(1)}(n-1)$ and by replacing $h_i(n)$ by $h'_i(n)$ for $i \in \{0,1\}$(Bayram & Selesnick, 2008).

DT-CWPT utilises dual-tree decomposition; thus, producing complex (real and imaginary) coefficients using dual-tree wavelet filters to the full binary tree. For j levels of decomposition, the wavelet packets produce 2^j different sets of coefficients. At level 5, both wavelet packet filters will generate a total of 64 sub-bands ($2^5 \times 2$). A matrix size of

64 × *M* composed of wavelet packet coefficient (sub-bands × coefficients) obtained, as described in Equation 1 below.

$$A = \left[C_5^1(M)\ \ C_5^2(M) \ldots\ C_5^{64}(M)\ \right]^T \quad [1]$$

To investigate the influence of wavelet levels, experiments using all levels from 1 to 5 ($[2^1 + 2^2 + 2^3 + 2^4 + 2^5] \times 2$), which produced 124 sub-bands were also conducted. The results just give a little difference in ±1%–2% accuracy with a longer computation time, so only the fifth level of DT-CWPT was chosen.

The Energy, Shannon and Renyi entropy measures were applied to the decomposed fifth level DT-CWPT sub-bands to extract a useful and straightforward feature vector. Those non-linear entropies were extracted to evaluate the subtle changes present in analysing non-stationary signals like speech signals and various bio-signals (Hariharan et al., 2018). The energy (EGY) of each wavelet packet sub-band coefficients was computed using Equation 2 below:

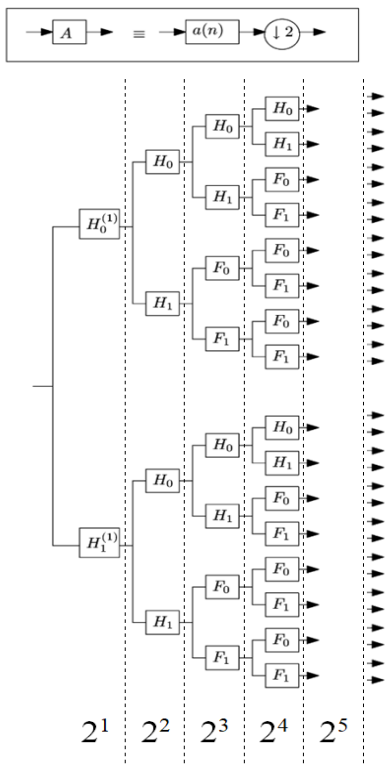

Figure 2. First wavelet packet FB of a five-level DT-CWPT. Note that the same decomposition mechanism also applies to the second wavelet FB.

$$EGY = \sum_{j,k} |C_{j,k}|^2 \quad [2]$$

The Shannon entropy defined by Equation 3, is an average information content measure that has been hidden in a signal. It's exploited to model the unpredictability and irregularities of a pathologic speech signal, as well as the possibility within a certain wavelet packet decomposition sub-band.

$$ShEn = -\sum_{j,k} C_{j,k} \log\left(C_{j,k}\right) \quad [3]$$

Renyi entropy is a well-known one-parameter generalisation of Shannon entropy. It is used to estimate the spectral complexity of a time series signal given by Equation 4 where α≠1.

$$ReEn = \frac{1}{1-\alpha} \log\left(\sum_{j,k} C_{j,k}^{\alpha}\right) \quad [4]$$

j=1, 2, 3,…*j*, where *j* is the number of decomposition level and *k*=1, 2, 3,…*N* is the number of wavelet packet coefficient in the respective sub-band.

The proposed method was then compared to the standard and well-known feature extraction methods, i.e. the MFCC and LPC (Alim & Rashid, 2018; Ankışhan, 2018). MFCC and LPC methods transform the voice signal from time-based to frequency-based domain while the DT-CWPT provides a time-frequency analysis of the voice signals. The 13 MFCC is chosen due to higher-order coefficients that represent increasing levels of spectral details; whereby depending on the sampling rate and estimation method, 12 to 20 cepstral coefficients are typically optimal for speech analysis (Huang et al.,2001; Virtanen et al., 2012). The order of ten for LPC is usually chosen since there is no significant improvement in sound quality for orders greater than ten (Ngo & Mehrubeoglu, 2010).

Classification

Two common classifiers, k-NN and SVM, were used to find the classification rate. The two-class analysis produced a result of normal or abnormal voice, while the multiclass analysis produced results according to the pathological voice. The k-NN classifier was chosen due to its simple implementation and flexibility to feature or distance choices (Abdullah et al., 2019). The classification was based on the majority of the k-Nearest Neighbor's category. In this study, k values were varied between 1 and 10. Instead of modelling the probability density of each class, SVM models the boundary between the classes. In biomedical applications, it is better to get a false alarm than a false negative, and the SVM seems to have better behaviour (Godino-Llorente et al., 2005). The best combination of two SVM parameters; cost (c) and gamma (γ) were obtained using LIBSVM Selection Tool (Chang & Lin, 2011). SVM was chosen since it has a better generalisation (less overfitting) and robust to noise.

In this work, a 10-fold cross-validation classification (CVC) scheme was used to increase the reliability of the results. Using this scheme, the extracted features were distributed into ten sets randomly, and ten times repetitive training was performed. To evaluate the two-class classifier performance, measures from the confusion matrix represented in Table 2 are considered. True positive (TP) measure of the classifier is classified as pathology when pathological samples are present, true negative (TN) classified as normal when normal samples are present, false positive (FP) classified as pathological when normal samples are present, and false negative (FN) classified as normal when pathological samples are present.

Table 2
Two-class confusion matrix

System decision	Actual diagnosis	
	Pathological	Normal
Pathological	True positive (TP)	False positive (FP)
Normal	False negative (FN)	True negative (TN)

The overall accuracy is calculated using the measures in Equation 5.

$$\text{Accuracy} = ((TP+TN)/ (\text{Total Samples})) \times 100\ \% \quad [5]$$

The performance of the multiclass analysis was evaluated by a confusion matrix represented in Table 3; where n = number of class. This matrix shows which points are correctly classified and which points are incorrectly classified. The number of test instances is shown by each matrix element for which the actual class is the row, and the predicted class is the column. Large numbers down the diagonal and small values (ideally zero) in the rest of the matrix relate to promising results.

Table 3
Multiclass confusion matrix

		Prediction			
		Class 1	Class 2	…	Class *n*
Actual	Class 1	Accuracy 1			
	Class 2		Accuracy 2		
	…			…	
	Class *n*				Accuracy *n*

The overall performance of the classifier is calculated as in Equation 6.

$$\text{Overall Accuracy} = \frac{(\text{Accuracy 1} + \text{Accuracy 2} + \ldots + \text{Accuracy n})}{\text{Total Samples}} \times 100\ \% \quad [6]$$

RESULT AND DISCUSSION

Overall, in this proposed work, the DT-CWPT based on Energy, Shannon and Renyi entropy, tested with k-NN and SVM classifiers yielded promising results. The results achieved better accuracy in the framed-based approach for all five datasets of voice samples compared to file-based analysis. Table 4 shows the two-class analysis for MEEI database (Dataset 1 and Dataset 2) and SVD database (Dataset 3). The proposed method, DT-CWPT with Shannon entropy, achieved the accuracy of 99.60% and 99.43% for Dataset 1 and 2 respectively while for Dataset 3, 94.60% obtained from DT-CWPT with Renyi entropy. From the Table 4, in the file-based approach the results outperformed the other two conventional methods (the highest accuracy of MFCC and LPC are 94.04% and 90.01% respectively). The frame-based experiment also gave good performance using the proposed method. The best performance was 100% accuracy score, achieved for both k-NN and SVM classifier in both Dataset 1 and 2, while the Dataset 3, best performance gave about 99.92% for k-NN and 99.94% for SVM.

Table 5 compares the proposed work with previous related researches for two-class analysis. The related works in Table 5 were selected because they used the same database and similar classifier as the proposed work. The difference in the feature extraction method employed would be an ideal opportunity to compare our results with those present in the

Table 4

Two-class analysis for the databases

2 CLASS	Classifier	Feature Extraction Method (no. of Coefficients)		Dataset 1 Accuracy (%) ± sd	Dataset 2 Accuracy (%) ± sd	Dataset 3 Accuracy (%) ± sd
FILE - BASED	KNN	DTCWPT (64)	Energy Entropy	88.54 ± 0.76	89.31 ± 1.22	81.34 ± 0.15
			Shannon Entropy	99.60 ± 0.14	99.43 ± 0.20	81.45 ± 0.35
			Renyi Entropy	92.83 ± 0.72	93.33 ± 0.80	82.41 ± 0.34
		MFCC (13)		82.12 ± 1.18	81.01 ± 1.51	84.61 ± 0.26
		LPC (10)		84.16 ± 0.75	83.08 ± 1.00	82.07 ± 0.40
	SVM	DTCWPT (64)	Energy Entropy	90.27 ± 0.81	88.93 ± 0.44	94.40 ± 0.30
			Shannon Entropy	99.20 ± 0.35	99.43 ± 0.20	90.01 ± 0.45
			Renyi Entropy	94.29 ± 0.64	91.76 ± 0.46	94.60 ± 0.29
		MFCC (13)		86.19 ± 0.85	85.28 ± 1.19	94.04 ± 0.35
		LPC (10)		87.21 ± 0.82	85.22 ± 0.68	90.01 ± 0.46
FRAME - BASED	KNN	DTCWPT (64)	Energy Entropy	100 ± 0.00	100 ± 0.00	99.88 ± 0.02
			Shannon Entropy	100 ± 0.00	100 ± 0.00	99.92 ± 0.02
			Renyi Entropy	100 ± 0.00	100 ± 0.00	99.88 ± 0.01
		MFCC (13)		100 ± 0.00	100 ± 0.00	99.97 ± 0.01
		LPC (10)		98.11 ± 0.10	99.53 ± 0.05	95.18 ± 0.09

Table 4 *(Continued)*

2 CLASS	Classifier	Feature Extraction Method (no. of Coefficients)		Dataset 1 Accuracy (%) ± sd	Dataset 2 Accuracy (%) ± sd	Dataset 3 Accuracy (%) ± sd
FRAME - BASED	SVM	DTCWPT (64)	Energy Entropy	99.99 ± 0.01	100 ± 0.00	99.94 ± 0.01
			Shannon Entropy	99.99 ± 0.03	99.97 ± 0.03	99.59 ± 0.03
			Renyi Entropy	100 ± 0.01	100 ± 0.00	99.94 ± 0.01
		MFCC (13)		99.99 ± 0.01	100 ± 0.00	99.92 ± 0.01
		LPC (10)		98.51 ± 0.11	99.73 ± 0.07	97.23 ± 0.05

Table 5

Overview of two-class analysis using MEEI subset (53 normal and 173 pathological)

Method	Feature	Classifier	Accuracy (%) ± sd	
			File-based	Frame-based
(Godino-Llorente et al., 2005)	MFCC with noise features	SVM	95.00 ± 2.00	94.10 ± 2.00
(Hariharan et al., 2014)	WPT	k-NN	99.65 ± 0.19	94.05 ± 0.83
		LS-SVM	99.12 ± 0.47	95.25 ± 0.12
(Majidnezhad (2015)	WPD with MFCC	ANN	94.24	-
(Muhammad et al., 2017)	IDP	SVM	99.38	-
Proposed	DT-CWPT	k-NN	99.60 ± 0.14	100 ± 0.00
		SVM	99.20 ± 0.35	100 ± 0.01

literature. Moreover, they had also analysed the data using a frame-based and file-based analysis, which is similar to the proposed work.

In the frame-based analysis, the proposed method demonstrated improvements in performance because more information in time and frequency scale was obtained from 5^{th} level proposed complex coefficients ($2\times2^5=64$) as compared to the 5^{th} level WPT coefficients

(2^5=32) proposed by Hariharan et al. (2014), thus generating real and imaginary tree fine resolution frequency sub-band data allowing for a better analysis. Moreover, a higher level of wavelet packet decomposition leads to better discriminative quality (Akbari & Arjmandi, 2014). These points contribute to better results performance compared to previous work.

These two-class accuracy results motivate the investigation of multiclass file-based and frame-based experiments for the other multiclass datasets in the databases. The complete results for multiclass analysis are shown in Table 6 and Table 7 for Dataset 4 and Dataset 5, respectively. The performance of the multiclass results was found not consistent. It is because of a limited number of pathology available and unevenly distributed number of samples from a different set of pathological voice in these databases. These limitations in the number and sample differences contribute to the accuracy performance for both databases.

It is known that the classification accuracy of vocal fold pathology detection systems is extremely dependent on the dataset and its characteristics, such as the volume of the dataset (Majidnezhad, 2015). Therefore, an adaptive synthetic (ADASYN) sampling approach is applied to imbalanced experimental datasets, to balance up the minority sample data to achieve better accuracy. ADASYN generates a weighted distribution for different minority class examples according to their level of difficulty in learning, where more synthetic data is produced for minority class examples that are harder to learn compared to those minority examples that are easier to learn. As a result, the ADASYN approach improves learning to the data distributions in two ways: reducing the bias introduced by the class imbalance, and adaptively shifting the classification decision boundary toward the difficult examples (He et al., 2008).

The proposed frame-based multiclass analysis using DT-CWPT with entropy and SVM yields a better average result ranging from 99.48%–99.65% as compared to 94.09%–98.80% obtained from its file-based analysis. However, the research on multiclass pathology analysis is lacking. The same database and almost similar datasets used by Muhammad et al. (2017) were applied in this work as a fair comparison, except for Class 1, where the authors used Cyst pathology for both MEEI and SVD database. The performances of file-based multiclass analysis of the proposed method are comparable, as indicated in Table 8.

Generally, both two-class and multiclass pathology detection using proposed DT-CWPT, produced better accuracy in the frame-based compared to the file-based analysis. This is because the frame based method framed the signal at 40ms per frame, which gives better time resolution analysis. It is known that speech signal exhibit quasi-stationary behaviour within the short period of time. In order to reduce feature loss and increase the continuity between adjacent frames in the framing, each frame is multiplied by Hamming window with 50% overlapped.

Nonetheless, the drawback of the frame-based analysis is that it takes a longer processing time due to more information obtained from all of the frames. The proposed feature methods also exhibit a small standard deviation (sd) showing the result ranges are more precise and give better performance using frame-based in both two-class and multiclass analyses.

Table 6

Three class classification for dataset 4

Dataset 4 (MEEI)	Classifier	Feature Extraction Method (no. of Coefficients)		Average Accuracy (%) ± sd	Class 1 Accuracy (%) ± sd	Class 2 Accuracy (%) ± sd	Class 3 Accuracy (%) ± sd
FILE - BASED (adasyn)	KNN	DTCWPT (64)	Energy Entropy	78.21 ± 0.77	95.97 ± 1.01	42.09 ± 1.97	96.57 ± 0.72
			Shannon Entropy	71.47 ± 1.12	92.19 ± 1.80	29.40 ± 2.64	94.09 ± 1.12
			Renyi Entropy	77.44 ± 1.18	97.10 ± 0.00	39.10 ± 2.88	95.52 ± 2.63
		MFCC (13)		77.15 ± 0.58	94.93 ± 1.02	44.93 ± 1.64	91.72 ± 1.96
		LPC (10)		71.96 ± 0.92	94.06 ± 1.27	31.04 ± 2.52	89.85 ± 1.46
	SVM	DTCWPT (64)	Energy Entropy	93.38 ± 0.91	99.40 ± 1.26	90.30 ± 1.27	90.45 ± 2.01
			Shannon Entropy	87.26 ± 1.56	86.09 ± 1.72	93.28 ± 1.27	82.27 ± 2.77
			Renyi Entropy	94.09 ± 0.99	94.35 ± 1.99	97.01 ± 0.00	90.90 ± 3.18
		MFCC (13)		90.75 ± 1.01	92.03 ± 1.23	95.07 ± 1.01	84.84 ± 2.95
		LPC (10)		83.82 ± 1.33	78.55 ± 2.54	95.67 ± 0.47	77.50 ± 2.20

Table 6 *(Continued)*

Dataset 4 (MEEI)	Classifier	Feature Extraction Method (no. of Coefficients)		Average Accuracy (%) ± sd	Class 1 Accuracy (%) ± sd	Class 2 Accuracy (%) ± sd	Class 3 Accuracy (%) ± sd
FRAME – BASED (adasyn)	KNN	DTCWPT (64)	Energy Entropy	99.42 ± 0.05	98.51 ± 0.00	99.80 ± 0.13	100 ± 0.00
			Shannon Entropy	99.34 ± 0.06	98.51 ± 0.00	99.59 ± 0.16	100 ± 0.00
			Renyi Entropy	99.43 ± 0.06	98.51 ± 0.00	99.98 ± 0.05	99.85 ± 0.19
		MFCC (13)		99.42 ± 0.02	98.51 ± 0.00	99.82 ± 0.05	100 ± 0.00
		LPC (10)		97.74 ± 0.09	98.62 ± 0.18	96.58 ± 0.29	98.02 ± 0.13
	SVM	DTCWPT (64)	Energy Entropy	99.44 ± 0.07	98.51 ± 0.00	100 ± 0.00	99.87 ± 0.21
			Shannon Entropy	99.48 ± 0.02	98.51 ± 0.00	99.98 ± 0.05	100 ± 0.00
			Renyi Entropy	99.47 ± 0.04	98.47 ± 0.10	100 ± 0.00	100 ± 0.00
		MFCC (13)		99.48 ± 0.00	98.51 ± 0.00	100 ± 0.00	100 ± 0.00
		LPC (10)		98.53 ± 0.12	98.47 ± 0.23	99.12 ± 0.19	98.02 ± 0.20

Table 7

Three class classification for dataset 5

Dataset 5 (SVD)	Classifier	Feature Extraction Method (no. of Coefficients)		Average Accuracy (%) ± sd	Class 1 Accuracy (%) ± sd	Class 2 Accuracy (%) ± sd	Class 3 Accuracy (%) ± sd
FILE - BASED (adasyn)	KNN	DTCWPT (64)	Energy Entropy	83.52 ± 0.50	100 ± 0.00	51.70 ± 1.48	98.03 ± 0.40
			Shannon Entropy	84.34 ± 0.30	100 ± 0.00	53.87 ± 0.70	98.42 ± 0.39
			Renyi Entropy	85.23 ± 0.38	100 ± 0.00	56.44 ± 1.12	98.75 ± 0.35
		MFCC (13)		87.38 ± 0.51	100 ± 0.00	64.74 ± 1.38	97.10 ± 0.39
		LPC (10)		85.33 ± 0.66	100 ± 0.00	61.49 ± 1.65	94.95 ± 0.76
	SVM	DTCWPT (64)	Energy Entropy	98.80 ± 0.26	100 ± 0.00	96.80 ± 0.59	99.56 ± 0.43
			Shannon Entropy	94.95 ± 0.61	100 ± 0.00	90.41 ± 0.81	94.43 ± 1.16
			Renyi Entropy	98.32 ± 0.39	100 ± 0.00	96.96 ± 0.75	98.00 ± 0.78
		MFCC (13)		96.77 ± 0.32	100 ± 0.00	97.58 ± 0.25	92.85 ± 0.82
		LPC (10)		93.36 ± 0.41	100 ± 0.00	90.05 ± 0.73	90.00 ± 1.60

Table 6 *(Continued)*

Dataset 5 (SVD)	Classifier	Feature Extraction Method (no. of Coefficients)		Average Accuracy (%) ± sd	Class 1 Accuracy (%) ± sd	Class 2 Accuracy (%) ± sd	Class 3 Accuracy (%) ± sd
FRAME – BASED (adasyn)	KNN	DTCWPT (64)	Energy Entropy	99.59 ± 0.01	100 ± 0.00	98.78 ± 0.03	100 ± 0.00
			Shannon Entropy	99.53 ± 0.03	100 ± 0.00	98.62 ± 0.06	99.98 ± 0.03
			Renyi Entropy	99.58 ± 0.03	100 ± 0.00	98.75 ± 0.09	100 ± 0.00
		MFCC (13)		99.63 ± 0.01	100 ± 0.00	98.89 ± 0.03	100 ± 0.00
		LPC (10)		97.39 ± 0.08	100 ± 0.00	94.06 ± 0.15	98.08 ± 0.14
	SVM	DTCWPT (64)	Energy Entropy	99.65 ± 0.01	100 ± 0.00	98.97 ± 0.00	99.99 ± 0.02
			Shannon Entropy	99.61 ± 0.02	99.98 ± 0.03	98.93 ± 0.06	99.93 ± 0.03
			Renyi Entropy	99.65 ± 0.01	99.99 ± 0.02	98.97 ± 0.00	99.99 ± 0.02
		MFCC (13)		99.64 ± 0.00	99.94 ± 0.00	98.97 ± 0.00	100 ± 0.00
		LPC (10)		99.04 ± 0.05	99.94 ± 0.00	97.57 ± 0.10	99.57 ± 0.09

Table 8

Accuracy of the methods for multiclass analysis (file-based)

Database / Method		Accuracy (%)			
		Class 1	Class 2	Class 3	Average
MEEI	Proposed DT-CWPT	94.35	97.01	90.90	94.09
	IDP (Muhammad et al., 2017)	99.10	94.30	94.50	95.97
SVD	Proposed DT-CWPT	100.00	96.80	99.56	98.80
	IDP (Muhammad et al., 2017)	99.50	95.90	95.10	96.83

Note: that the proposed class definition is as defined in Table 1

CONCLUSION

This work investigated feature extraction based on the DT-CWPT using energy and entropy measures tested with two classifiers, k-NN and SVM. The DT-CWPT performance as a feature extraction tool was proven to be reliable to detect the presence of diseases of the vocal fold. The proposed features yielded promising results and surpassed the conventional MFCC and LPC performance for file-based approach. A new set of features (real and imaginary coefficients) from the signal decomposition contribute to produce the best overall performance in detecting specific pathologies. The proposed system can be used to discriminate between two-class (normal and abnormal) and multiclass samples of voice pathologies. The experimental results using the proposed DT-CWPT features for the two-class analysis achieved 100% and 99.94% accuracy for MEEI and SVD database, respectively. Meanwhile, 99.48% for MEEI database and 99.65% for SVD database were achieved in multiclass. In future, it is hoped that more pathological samples can be obtained from these databases and also from other available databases so that more other specific pathology can be diagnosed and use worldwide. Feature optimisation can also be employed to further optimise the features obtained from DT-CWPT. It may include feature reduction and feature selection optimisation method.

ACKNOWLEDGEMENT

This work was done in part with data from the SVD database: http://www.stimmdatenbank.coli.uni-saarland.de/. The authors would like to thank the anonymous reviewers for their valuable comments.

REFERENCES

Abdullah, R., Muthusamy, H., Vijean, V., Abdullah, Z., & Kassim, F. N. C. (2019). Real and complex wavelet transform approaches for Malaysian speaker and accent recognition. *Pertanika Journal of Science and Technology*, *27*(2), 737-752.

Akbari, A., & Arjmandi, M. K. (2014). An efficient voice pathology classification scheme based on applying multi-layer linear discriminant analysis to wavelet packet-based features. *Biomedical Signal Processing and Control*, *10*(1), 209-223.

Alim, S. A., & Rashid, N. K. A. (2018). Some commonly used speech feature extraction algorithms. In R. Lopez-Ruiz (Ed.), *From natural to artificial intelligence-algorithms and applications* (pp. 4-22). London, UK: IntechOpen.

Ankışhan, H. (2018). A new approach for detection of pathological voice disorders with reduced parameters. *Electrica*, *18*(1), 60-71.

Barry, W. J., & Pützer, M. (2007). *Saarbruecken voice database*. Institute of Phonetics, University of Saarland. Retrieved July 9, 2018, from http://www.stimmdatenbank.coli.uni-saarland.de/

Bayram, I., & Selesnick, I. W. (2008). On the dual-tree complex wavelet packet and M-band transforms. *IEEE Transactions on Signal Processing*, *56*(6), 2298-2310.

Cao, X. C., Chen, B. Q., Yao, B., & He, W. P. (2019). Combining translation-invariant wavelet frames and convolutional neural network for intelligent tool wear state identification. *Computers in Industry*, *106*, 71-84.

Chang, C. C., & Lin, C. J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *2*(3), 1-27.

Godino-Llorente, J. I., Gomez-Vilda, P., Sáenz-Lechón, N., Blanco-Velasco, M., Cruz-Roldán, F., & Ferrer-Ballester, M. A. (2005). Support vector machines applied to the detection of voice disorders. *Lecture Notes in Computer Science*, *3817*, 219-230.

Haidong, S., Junsheng, C., Hongkai, J., Yu, Y., & Zhantao, W. (2019). Enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy for early fault prognosis of bearing. *Knowledge-Based Systems, 188*, 1-14.

Harar, P., Galaz, Z., Alonso-Hernandez, J. B., Mekyska, J., Burget, R., & Smekal, Z. (2018). Towards robust voice pathology detection: Investigation of supervised deep learning, gradient boosting, and anomaly detection approaches across four databases. *Neural Computing and Applications*, *7*, 1-11.

Hariharan, M., Polat, K., & Yaacob, S. (2014). A new feature constituting approach to detection of vocal fold pathology. *International Journal of Systems Science*, *45*(8), 1622-1634.

Hariharan, M., Sindhu, R., Vijean, V., Yazid, H., Nadarajaw, T., Yaacob, S., & Polat, K. (2018). Improved binary dragonfly optimization algorithm and wavelet packet based non- linear features for infant cry classification. *Computer Methods and Programs in Biomedicine*, *155*(December), 39-51.

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *Proceedings of the International Joint Conference on Neural Networks*, *3*, 1322-1328.

Huang, X., Acero, A., Hon, H. W., & Foreword By-Reddy, R. (2001). *Spoken language processing: A guide to theory, algorithm, and system development*. Upper Saddle River, United States: Prentice Hall PTR.

Lim, W. J., Muthusamy, H., Vijean, V., Yazid, H., Nadarajaw, T., & Yaacob, S. (2018). Dual-tree complex wavelet packet transform and feature selection techniques for infant cry classification. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *10*(1-16), 75-79.

Majidnezhad, V. (2015). A novel hybrid of genetic algorithm and ANN for developing a high efficient method for vocal fold pathology diagnosis. *Eurasip Journal on Audio, Speech, and Music Processing*, *1*, 1-11.

Martinez, D., Lleida, E., Ortega, A., Miguel, A., & Villalba, J. (2012). Voice pathology detection on the Saarbrücken Voice Database with calibration and fusion of scores using multifocal toolkit. *Communications in Computer and Information Science*, *328*, 99-109.

Mekyska, J., Janousova, E., Gomez-Vilda, P., Smekal, Z., Rektorova, I., Eliasova, I., … & López-de-Ipiña, K. (2015). Robust and complex approach of pathological speech signal analysis. *Neurocomputing*, *167*, 94-111.

Muhammad, G., Alsulaiman, M., Ali, Z., Mesallam, T. A., Farahat, M., Malki, K. H., … & Bencherif, M. A. (2017). Voice pathology detection using interlaced derivative pattern on glottal source excitation. *Biomedical Signal Processing and Control*, *31*, 156-164.

Ngo, H., & Mehrubeoglu, M. (2010). Effect of the number of LPC coefficients on the quality of synthesized sounds. *International Journal of Engineering Research and Innovation, 2*(2), 11-16.

Patil, H. A. (2019). Combining evidences from variable teager energy source and mel cepstral features for classification of normal vs. pathological voices. *European Signal Processing Conference*, *2019-September*(2), 1-5.

Qu, J., Zhang, Z., & Gong, T. (2016). A novel intelligent method for mechanical fault diagnosis based on Dual-tree Complex Wavelet Packet Transform and multiple classifier fusion. *Neurocomputing*, *171*, 837-853.

Saidi, P., & Almasganj, F. (2015). Voice disorder signal classification using M-Band wavelets and support vector machine. *Circuits, Systems, and Signal Processing*, *34*(8), 2727-2738.

Selesnick, I. W., Baraniuk, R. G., & Kingsbury, N. C. (2005). The Dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, *22*(6), 123-151.

Serbes, G., Aydin, N., & Gulcur, H. O. (2013). Directional dual-tree complex wavelet packet transform. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2211*, 3046-3049.

Shafik, A., Elhalafawy, S. M., Diab, S. M., Sallam, B. M., & Abd El-samie, F. E. (2009). A wavelet based approach for speaker identification from degraded speech. *International Journal of Communication Networks and Information Security*, *1*(3), 52-58.

Srinivasan, V., Ramalingam, V., & Arulmozhi, P. (2014). Artificial neural network based pathological voice classification using MFCC features. *International Journal of Science, Environment and Technology*, *3*(1), 291-302.

Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal acoustic analysis–jitter, shimmer and hnr parameters. *Procedia Technology*, *9*, 1112-1122.

Vikram, C. M., & Umarani, K. (2013). Pathological voice analysis to detect neurological disorders using MFCC and SVM. *International Journal of Advanced Electrical and Electronics Engineering, (IJAEEE)*, *4*, 87-91.

Virtanen, T., Singh, R., & Raj, B. (Eds.). (2012). *Techniques for noise robustness in automatic speech recognition*. Chichester, UK: John Wiley & Sons.