

OPTIMIZATION OF DUAL-TREE COMPLEX WAVELET PACKET BASED ENTROPY FEATURES FOR VOICE PATHOLOGIES DETECTION

Farah Nazlia Che Kassim^{a*}, Vikneswaran Vijejan^a, Zulkapli Abdullah^a, Hariharan Muthusamy^b, Rokiah Abdullah^a

^aFaculty of Electronic Engineering Technology, Universiti Malaysia Perlis (UniMAP), 02600 Arau, Perlis, Malaysia

^bDepartment of Electronic Engineering, National Institute of Technology, Srinagar (Garhwal), Uttarakhand India

Article history

Received

11 March 2020

Received in revised form

1 October 2020

Accepted

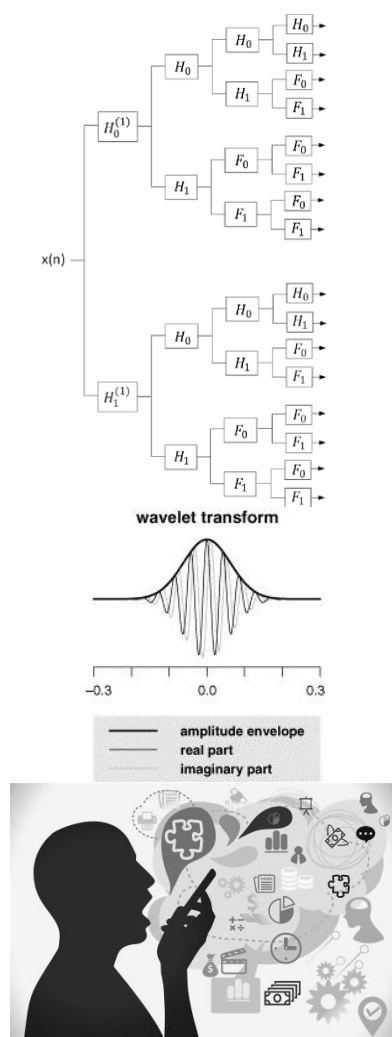
5 October 2020

Published online

27 October 2020

*Corresponding author
nazlia@unimap.edu.my

Graphical abstract



Abstract

The Dual-Tree Complex Wavelet Packet Transform (DT-CWPT) has been successfully implemented in numerous fields because it introduces limited redundancy, provides approximate shift-invariance and geometrically oriented signal in multiple dimensions where these properties are lacking in traditional wavelet transform. This paper investigates the performance of features extracted using DT-CWPT algorithms which are quantified using k -Nearest Neighbors (k -NN) and Support Vector Machine (SVM) classifiers for detecting voice pathologies. Decomposition is done on the voice signals using Shannon and Approximate entropy (ApEn) to signify the complexity of voice signals in time and frequency domain. Feature selection methods using the Relief algorithm and Genetic algorithm (GA) are applied to obtain the optimum features for multiclass classification. It is observed that the best accuracies obtained using DT-CWPT with ApEn entropy are 91.15% for k -NN and 93.90% for SVM classifiers. The proposed work provides a promising detection rate for multiple voice disorders and is useful for the development of computer-based diagnostic tools for voice pathology screening in health care facilities.

Keywords: DT-CWPT, Feature Extraction, Feature Selection, Voice Pathologies

Abstrak

Kaedah *Dual-Tree Complex Wavelet Transform* (DT-CWPT) telah berjaya dilaksanakan dalam pelbagai bidang kerana ia menghadkan kelebihan, menyediakan anggaran peralihan anjakan dan isyarat berorientasikan geometri pelbagai dimensi di mana ciri-ciri ini kurang dalam transformasi wavelet yang tradisional. Artikel ini mengkaji prestasi kaedah pengekstrakan ciri dengan menggunakan algoritma DT-CWPT yang diklasifikasi menggunakan teknik k -Nearest Neighbors (k -NN) dan *Support Vector Machine* (SVM) untuk mengesan pathologi suara. Penghuraian komposisi suara dilakukan dengan menggunakan entropi Shannon dan Anggaran (ApEn) untuk mengurangkan kerumitan isyarat suara dalam domain masa dan frekuensi. Kaedah pemilihan ciri menggunakan algoritma Relief dan genetik (GA) digunakan untuk mendapatkan ciri-ciri optimum pelbagai kelas. Pengelasan terbaik yang diperoleh menggunakan DT-CWPT dengan entropi ApEn adalah 91.15% untuk k -NN dan 93.90% untuk SVM. Kaedah yang dicadangkan ini memberikan kadar pengesanan pelbagai penyakit gangguan suara dan berguna untuk pembangunan alat diagnostik berasaskan komputer untuk pemeriksaan patologi suara di kemudahan penjagaan kesihatan.

Kata kunci: DT-CWPT, Pengekstrakan Ciri, Pemilihan Ciri, Pathologi Suara

© 2020 Penerbit UTM Press. All rights reserved

1.0 INTRODUCTION

Feature extraction is a conversion sequence of speech samples process into observation vectors set which represent events in a probabilistic space over which classification is performed [1]. Numerous literatures have been published in voice pathology fields using acoustic parameters such as pitch, formants, jitter and shimmer, Noise-to-Harmonic Ratio, Soft Phonation Index, Relative Average Perturbation, glottal-to-noise excitation ratio, Amplitude Perturbation Quotient, Mel-Frequency Cepstral Coefficients (MFCC), modulation spectra, bicepstrum, inferior colliculus coefficients, sample and ApEn entropy and empirical mode decomposition [2, 3 and 4]. Recently, significant attention has been addressed to analyze signals in time-frequency domain analysis namely wavelet analysis. Wavelets provide a theory analogous to Fourier analysis and have been used for data compression in both image and sound processing applications. Since the voice signal is a highly non-stationary signal, Fourier transform is not a very suitable tool to analyze non-stationary signal as the time-domain information is lost during the frequency transformation [5]. Since conventional wavelets based features have several drawbacks such as ringing, shift variance, aliasing and lack of directionality [6], wavelets have been evolved. Hariharan *et al.* [5] proposed the Wavelet Packet Transform (WPT) and singular value decomposition for normal or pathological vocal fold pathology detection. The proposed weighted features using k-means clustering-based gives a promising result of almost 100% accuracy. This work involved 57 normal and 173 pathological signals from Massachusetts Eye and Ear Infirmary (MEEI) voice disorders database whilst 24 normal and 24 pathological voices from MAPACI speech pathology database. Saidi and Almasganj [7] classified the normal or pathological voice using a five-band wavelet system employing a GA to determine the optimal wavelet parameters and shown good performance close to 100%. Different samples containing 57 normal and 653 pathological signals from MEEI database were used. Wavelet Packet Decomposition (WPD) combined with MFCC as feature extraction introduced by Majidnezhad [8] developed a higher accuracy of 94.24% and lower response time with reduced feature vector from 139 to 30 features. This approached were based on the hybrid of GA as feature reduction and artificial neural network as the classifier using MEEI database.

A four-class classification of voice disorders has been studied by Akbari and Arjmandi [9] using Discrete Wavelet Packet Transform (DWPT) tested on the MEEI database with 53 normal voice samples, 72 A-P squeezing samples, 85 hyperfunction samples and 48 samples from gastric reflux. The use of Shannon entropy and energy features was able to provide the precision of 93.33% and 94.67%, respectively. These best rates achieved by level eighth of DWPT using Multi-class Linear Discriminant Analysis (MC-LDA), with 256 coefficients classified by the multi-layer neural network. Che Kassim *et al.* [10] reported the use of DT-CWPT based features gave

99.43% accuracy for pairwise classification tested using 53 normal and 106 abnormal samples from MEEI database. The abnormal samples were further studied in multiclass classification where the highest achievement reported from DT-CWPT with entropy features was 78.21% and 94.09% using k-NN and SVM classifier respectively. To overcome the low performances in multiclass analysis obtained from full 64 features produced by DT-CWPT, feature selection using ReliefF was proposed by the same authors [11]. However, a small increased in accuracy and reduced classification time with the selected 30 features acquired.

A binary classification deals with two classes (abnormal and normal) while multiclass handles more than two classes of vocal fold pathology which means a sample needs to be determined between more query instances, therefore problem becomes more complex and need more features. More features leads to higher computation as dimension of features are higher. Therefore, there is a need to reduce the feature dimension, either by feature reduction or selection methods that can reduce complexity and improved accuracy. Selection of salient features from the whole features is needed to reduce the misclassification error rate [8]. Some of the popular techniques for dimensional reduction are GA, Principal Component Analysis, ReliefF, Linear Discriminant Analysis, Higher-Order Singular Value Decomposition, Fisher Discrimination Ratio and Minimum Redundancy Maximum Relevance [12]. According to Roffo [13], the overfitting chances increase with the number of features, therefore, feature selection can improve algorithms performance and classification accuracy. These methods deal with selecting the best feature set and reducing correlation between features which speeds up training time and as well as improved model's performance to develop an efficient analysis tool. Extracting distinct quality features from pathologic voices appears to have the best outcome in the voice disorders classification [9].

Voice pathology database has a limited number of pathology available with unevenly distributed number of samples from different sets of pathological voice. These limitation number and sample differences contribute to the performances of multiclass classification. It is known that the classification accuracy of voice pathology detection systems extremely depends on the dataset and its characteristics such as the volume of dataset [8]. From the previous works discussed, it is observed that the wavelet packet method produce more features number from its subbands depends on chosen decomposition levels. DT-CWPT decomposition produces complex coefficients by using a dual-tree of wavelet filters to obtain real and imaginary coefficients. Thus, introduces good properties such as limited redundancy, provides approximately shift-invariance and geometrically oriented signal in multiple dimensions [14] which were lacking in conventional wavelet transform. Its multi-resolution feature extraction break a signal into sine waves of various frequencies capable of extracting local and global information revealing important aspects like trends, breakdown points,

discontinuities in higher derivatives, and self-similarity which characterizes a signal [15]. The transform is approximately shift-invariant meaning if the input sequence is changed by an arbitrary number of samples; the energy in each subband is preserved. The DT-CWPT is approximately analytic which results in the multi-dimensional form of the transform will be geometrically oriented in two or more dimensions.

With those properties mentioned, the complex coefficients obtained from the DT-CWPT decomposition give more advantages in good time-frequency resolution information. This would be useful as an effort to extract better quality features that can contribute to the overall best performance for detecting the multiple voice pathology. Also with features optimization, it demonstrates better performances as an efficient tool for vocal fold analysis. Hence, this motivates us to explore the DT-CWPT based features for multiple voice pathology detection with the optimized features from ReliefF and GA algorithms quantified using k -NN and SVM classifiers. The proposed work is expected to be useful for the development of computer-based diagnostic tools for voice pathology detection.

2.0 METHODOLOGY

Voice pathologies detection system was developed to distinguish between normal person and patient who have voice pathologies. The overall block diagram for the proposed voice pathology analysis is shown in Figure 1.

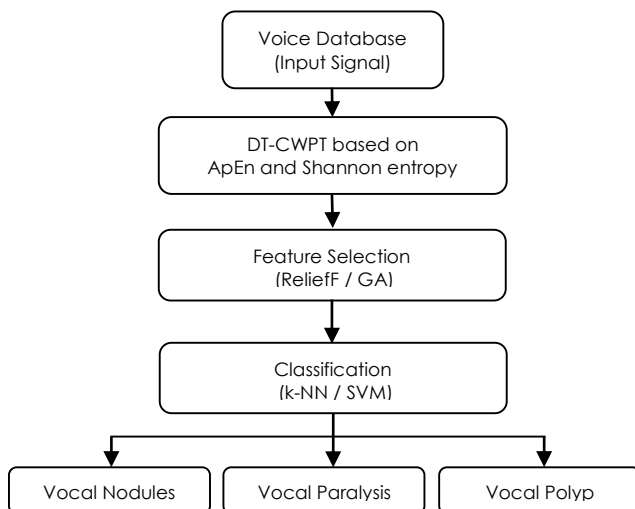


Figure 1 Proposed voice pathology analysis block diagram

MEEI voice disorder database is used for the experiments to evaluate the proposed method. MEEI database is the only commercially available database used in most studies of voice pathology detection and classification [16]. A total of 106 voice samples containing three types of pathological voice are used as shown in Table 1. It's difficult to compare the performance due to the different dataset and uneven numbers of samples chosen from different databases. Therefore, the data were selected as to standardize the samples and

benchmark with existing works in [10] and [11]. The voice samples are sustained vowel /a/ produced at normal pitch sampled with a 25 or 50 kHz sampling rate in a controlled environment. All files were downsampled to 25 kHz to have a uniform sampling frequency [2, 17].

Table 1 Voice samples used in multiclass pathology analysis

Pathology Class	MEEI Database	
1	Vocal Nodules	19
2	Vocal Paralysis	67
3	Vocal Polyp	20
Total Samples		106

2.1 Feature Extraction

The extraction of feature is to initially parameterize the voice signal to acquire the best-voiced sound representation for a specific class. WPT presents a time-frequency analysis based on quantifying cross-correlation between a mother wavelet and analyzed signal. The energy compaction and decorrelation properties of WPT, make the wavelet domain useful for signal processing especially in the parameterization of pathologic speech.

DT-CWPT is an extension to the Dual-Tree Complex Wavelet Transform (DT-CWT) algorithm. It is constructed by decomposing both sub-bands of first and second filter-bank repeatedly in the real and imaginary tree using low or high-pass perfect reconstruction filter-banks. The filter-banks used in the first decomposition should also be used to decompose the second filter-bank of the DT-CWT in order to preserve the Hilbert transform relationship of those branches. The first stage high-pass branch, $h_1^{(l)}(n)$ and $h_{-1}^{(l)}(n)$ satisfy $h_{-1}^{(l)}(n) = h_1^{(l)}(n-1)$. This is exactly the same relationship satisfied by the first stage low-pass filters, $h_0^{(l)}(n) = h_0^{(l)}(n-1)$. The second wavelet packet filter-bank is obtained by replacing the first stage filters $h_i^{(l)}(n)$ by $h_i^{(l)}(n-1)$ and by replacing $h_i(n)$ by $h_{-i}^{(l)}(n)$ for $i \in \{0, 1\}$. This process generates two wavelet packet filter-banks of DT-CWPT operating in parallel. Some filters are the same in the first and second wavelet packet filter-bank. Figure 2 below represented a four-level DT-CWPT for the first wavelet packet filter-bank. DT-CWPT possessed the same properties as DT-CWT such as shift-invariance and good directional selectivity but with fewer energy leakages into its negative frequency bands [14].

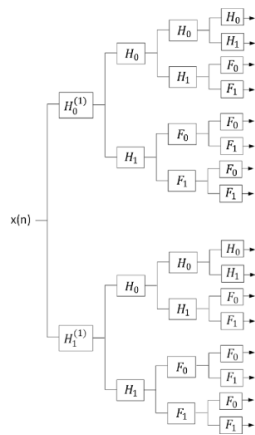


Figure 2 Four-level DT-CWPT for the first wavelet packet filter-bank [14]

The decomposition level employed to the wavelets determines the frequency intervals for decomposing the original signal. For j levels of decomposition, the wavelet packets produce 2^j sub-bands consisting of different coefficients sets. Several levels were investigated and it is observed that the best precision obtained from the fifth level of the wavelet decomposition. In this paper, the 5th level DT-CWPT yields 64 sub-bands ($2^5 \times 2$ wavelet packet filter-banks) and further calculated based on entropy measures, ApEn and Shannon entropy. The entropies were used to characterize the rate of creation of information from the extracted feature in the sub-bands and to signify the complexity of voice signal in both time and frequency domain.

Shannon entropy averaged the unpredictability and irregularities information that has been hidden in a pathologic speech signal within the WPD sub-bands. Shannon entropy is capable to model abnormalities appropriately in the vocal folds that result in local discontinuities, uncertain high random property series of events, quasi-noise components, signal sudden variations and distorted spectrum for analyzed speech [9]. The Shannon entropy is defined by the following equation (Equation 1).

$$ShEn = - \sum_{j,k} C_{j,k} \log(C_{j,k}) \quad (1)$$

Where

- j decomposition level number
- k wavelet packet coefficient number in the respective sub-band.

ApEn is a measure of regularity of data and complexity that evaluate efficiently even over relatively short time series which is well-suited for the pathological voice analysis. It measures the differences between the two time-series probabilities of m length templates that are close to each other and $m+1$ length templates that are close to each other. Therefore, ApEn shows the probability of creating a new pattern with increasing of the dimension template. The greater the probability values described more complex time series. ApEn introduced by Pincus *et al.* [18] is defined as in equation (Equation 2).

$$ApEn(m, r, N) = \Phi m(r) - \Phi m+1(r) \quad (2),$$

for an N -dimensional time series:

$$\Phi^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \ln C_i^m(r)$$

Where:

- \ln the natural logarithm
- m the embedding dimension
- r the threshold (filtering level)

2.2 Feature Selection

In this approach, the feature selection method using the ReliefF algorithm and GA was added to select the most relevant features while reducing redundancy for multiclass classification. Feature selection methods reduce the dimensions of features without much loss of the total information. The ReliefF is extended from Relief to support multiclass classification, supervised filter algorithms that select features to isolate instances from different classes. Euclidean distance is taken as correlation index and then features are weights according to how well they differentiate instances of different classes. ReliefF selects an instance (R_i) randomly then searches for k from the same class of its nearest neighbours, called nearest hits (H_j). k nearest neighbours from each of the different classes also been searched called nearest misses ($M_j(C)$). Depending on their values for R_i , H_j and $M_j(C)$, It updates the quality estimation $W[A]$ for all attributes A . Then it averages the contribution of all the hits and all the misses as in equation (Equation 3). Each class of the misses contribution is weighted with a prior probability of that class $P(C)$ which is predicted from the training set. Based on preliminary empirical testing, the value of $k=10$ was suggested and has been widely accepted as the default setting [19]. The best features ranking based on the weights of the features were then quantified by classifiers.

$$W[A] = W[A] - \sum_{j \in H} \frac{\text{diff}(A, R_i, H_j)}{(m, k)} + \sum_{C \in \text{class}(R_i)} \left[\frac{P(C)}{1 - P(\text{class}(R_i))} \sum_{j \in M} \text{diff}(A, R_i, M_j(C)) \right] / (m, k) \quad (3)$$

GA introduced by Holland [20] is inspired by Charles Darwin's natural evolution theory that reflects natural selection process where the fittest individuals are chosen for reproduction in order to generate next-generation offspring. It is a stochastic approach which studies environment nature and picks up the best solution among a potential solutions population [8, 21]. It considers as robust optimization and search methods that use some mechanisms such as initial population, natural selection, fitness function, reproduction, crossover, mutation and then chooses the optimal solution. The values of GA's parameters were adjusted experimentally to achieve a better performances where some parameters are also mentioned by Majidnezhad [8] and Oluleye *et al.* [22]. The suggested parameters used in this paper for the GA algorithm were 0.5 for the arithmetic crossover using a uniform mutation of 0.05 for 100 populations set as binary strings. A fitness function, *fit* in equation

(Equation 4) is defined to evaluate the features discriminative potential in each subset.

$$fit = \frac{\alpha}{N_f} + \exp\left(\frac{-1}{N_f}\right) \quad (4)$$

Where:

α the k -NN based classification error

N_f cardinality of the selected features

Lower fitness individuals have a better survival chance into the mating pool or next generation. The algebraic structure of this equation ensures the learning of the GA reduced number of features selected and minimized error [22].

2.3 Classification

The voice pathology classification can be seen as pattern recognition and are an important tool to discriminate between normal and pathological voice or to know the disease from a speech signal [23]. The selected features become the input to k -NN and SVM classifier aiming at identifying the three types of voice pathologies which are vocal nodules, paralysis and polyp in this study. A 10-fold cross-validation classification scheme is used to prove the reliability of the results for both classifiers. The data are divided equally into 10 segments where 10 repetitions of training and testing are performed in which; a different segment of data is used to avoid the overlap.

In k -NN, an object is classified by a majority vote of its neighbour, with the object being assigned to the class most common amongst its k -nearest neighbours. The number of neighbours, k is a positive integer varied between 1 and 10 for a given training and test data. The Euclidean distance measure is used to locate the nearest neighbour and find the closest members of the training set to the test class being examined. Equation 5 calculated the Euclidean distance metrics $d(x,y)$ between two points x and y .

$$d(x,y) = \sum_{i=1}^N \sqrt{x_i^2 - y_i^2} \quad (5)$$

The SVM builds a model, specifies the hyperplane that separates the different classes and maps the decision boundary for each class. The SVM with radial basis function as a kernel was utilized because it has less restriction on the data volume and number of features, more general than the linear kernels and it produces better accuracy compared to other kernel functions [24,25]. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier [26]. The best combination of two SVM parameters; cost (c) and gamma (γ) were obtained using LIBSVM selection tool which has been implemented by Chang and Lin [27].

3.0 RESULTS AND DISCUSSION

In this paper, several experiments were made by taking into account the number of features

produced and classification time taken before and after feature selection. Comparison of average multiclass performances using features produced by DT-CWPT based on ApEn and Shannon entropy with and without feature selection is shown in Figure 3. DT-CWPT produced 64 coefficients from 5th level and it can be seen that not all features are informative and useful. Overall results show the performance of features increased between 2 to 11% using GA with a reduced number of features. For k -NN classifier, DT-CWPT based on ApEn with GA gives the accuracy of 91.15 % and 87.51% for Shannon entropy. For each entropy mentioned earlier, a higher performance of 93.90% and 92.23% is obtained from SVM. The accuracy with ReliefF algorithm just gave a slight improvement about ± 1 -2% accuracy with reducing half of the DT-CWPT full features.

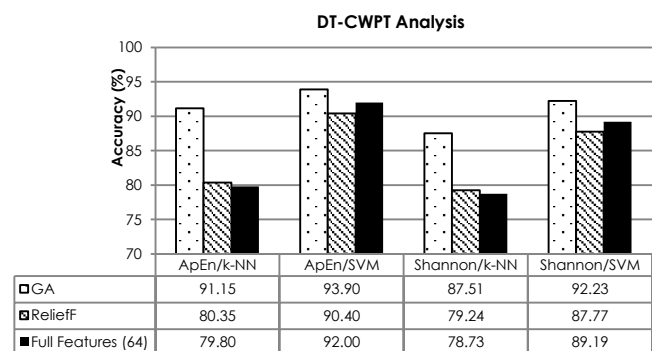


Figure 3 Performances of features produced by DT-CWPT before and after feature selection

Performances were comparable with Akbari and Arjmandi [9] using MC-LDA with wavelet packet-based energy and Shannon entropy which gave the accuracy of 93.33% and 94.67%, respectively. The work was selected as an ideal opportunity for comparison because they had used the same database with a conventional wavelet type. However, their work utilized a different size of the dataset, numbers of features extracted from a different type of voice pathologies using different classification methods. Further investigations were made and compared with 5th level WPT with the same feature selection and classifiers method. The details are as shown in Table 2 and Table 3 for each wavelet-based on ApEn using k -NN and SVM classifiers respectively. GA performed well in both wavelets as it increases the accuracy with reducing features to 7 out of 32 features for WPT and 18 features for DT-CWPT as well as decreased classification time.

Table 4 and Table 5 compare classification results of WPT and DT-CWPT based on Shannon entropy before and after feature selection using k -NN and SVM classifier. Again, GA performed well in both wavelets as it increases the accuracy with reducing features to 6 out of 32 features for WPT and 14 features for DT-CWPT. Overall, SVM shows better improvement in accuracy with reduced features and better classification time. Using all features it takes almost 20 seconds compared to after applying the feature reduction which takes only about 10

seconds while k-NN produced whether the same or a slightly improved in accuracy and time.

Table 2 Comparison of wavelet analysis based on ApEn entropy using k-NN classification

Wavelet Analysis	No. of Features	Accuracy (% ± sd)				Time (s)
		Average	Class 1	Class 2	Class 3	
WPT ApEn entropy						
GA	7	81.60 ± 1.05	90.62 ± 1.14	69.55 ± 1.89	84.80 ± 1.97	0.65
ReliefF	30	79.95 ± 0.83	92.00 ± 1.41	52.09 ± 1.31	95.88 ± 1.35	0.65
Full Features	32	80.15 ± 0.47	92.00 ± 1.41	52.84 ± 1.44	95.74 ± 0.83	0.69
DT-CWPT ApEn entropy						
GA	18	91.15 ± 0.71	95.29 ± 1.16	84.48 ± 1.60	93.69 ± 1.35	0.62
ReliefF	30	80.35 ± 1.63	95.15 ± 1.21	55.97 ± 2.56	90.00 ± 3.18	0.63
Full Features	64	79.80 ± 0.92	96.62 ± 0.99	51.04 ± 2.20	91.85 ± 1.27	0.71

Table 3 Comparison of wavelet analysis based on ApEn entropy using SVM classification

Wavelet Analysis	No. of Features	Accuracy (% ± sd)				Time (s)
		Average	Class 1	Class 2	Class 3	
WPT ApEn entropy						
GA	7	86.20 ± 1.16	91.69 ± 1.30	77.76 ± 2.48	89.26 ± 1.21	8.30
ReliefF	30	89.65 ± 0.82	94.15 ± 1.75	79.85 ± 2.14	95.00 ± 1.99	13.59
Full Features	32	91.20 ± 1.36	92.15 ± 1.98	88.66 ± 2.83	92.79 ± 1.76	13.61
DT-CWPT ApEn entropy						
GA	18	93.90 ± 0.61	95.88 ± 1.67	91.04 ± 0.00	94.77 ± 1.49	11.18
ReliefF	30	90.40 ± 0.57	97.94 ± 1.03	78.66 ± 1.58	94.62 ± 1.66	13.32
Full Features	64	92.00 ± 1.15	96.47 ± 1.24	89.55 ± 1.57	89.85 ± 2.63	18.58

Table 4 Comparison of wavelet analysis based on Shannon entropy using k-NN classification

Wavelet Analysis	No. of Features	Accuracy (% ± sd)				Time (s)
		Average	Class 1	Class 2	Class 3	
WPT Shannon entropy						
GA	6	81.11 ± 1.74	88.94 ± 1.44	68.06 ± 3.39	86.52 ± 2.52	0.61
ReliefF	30	79.40 ± 0.71	94.09 ± 0.86	52.09 ± 1.79	92.42 ± 1.43	0.63
Full Features	32	78.89 ± 0.95	94.70 ± 0.80	51.19 ± 2.34	91.21 ± 1.56	0.71
DT-CWPT Shannon entropy						
GA	14	87.51 ± 1.64	96.25 ± 1.68	72.84 ± 2.88	93.94 ± 1.24	0.56
ReliefF	30	79.24 ± 0.88	93.75 ± 1.65	48.51 ± 1.06	96.36 ± 1.28	0.62
Full Features	64	78.73 ± 1.00	94.84 ± 1.48	45.97 ± 1.37	96.36 ± 1.78	0.71

Table 5 Comparison of wavelet analysis based on Shannon entropy using SVM classification

Wavelet Analysis	No. of Features	Accuracy (% ± sd)				Time (s)
		Average	Class 1	Class 2	Class 3	
WPT Shannon entropy						
GA	6	85.38 ± 0.77	90.00 ± 1.63	80.00 ± 1.60	86.21 ± 2.31	9.02
ReliefF	30	85.53 ± 1.46	87.27 ± 2.87	82.09 ± 1.86	87.27 ± 2.28	14.23
Full Features	32	85.93 ± 1.52	87.88 ± 2.37	81.79 ± 1.37	88.18 ± 3.56	14.63
DT-CWPT Shannon entropy						
GA	14	92.23 ± 1.17	95.31 ± 2.76	90.45 ± 1.60	91.06 ± 1.81	10.96

ReliefF	30	87.77 ± 0.91	82.34 ± 1.96	95.07 ± 0.72	85.61 ± 1.79	14.38
Full Features	64	89.19 ± 1.35	84.38 ± 2.76	97.01 ± 0.00	85.91 ± 2.86	19.89

The best-achieved accuracy, 90.40% obtained from the combination of DT-CWPT based on ApEn and ReliefF (30 out of 64 features) using SVM. From observation for both wavelet methods, reducing features after the ReliefF algorithm gave the advantages in classification time (reduced 1-5s) with a little change in performances. As compared to GA, ReliefF method has a low computational burden in selection techniques because it evaluates and selects the feature subsets by focusing on the general characteristics without the chosen learning algorithm or the classifier been involved. It utilizes correlation and information techniques to search for the most pertaining features sets. ReliefF will not discriminate among the existing redundant features as it assembles the features by importance.

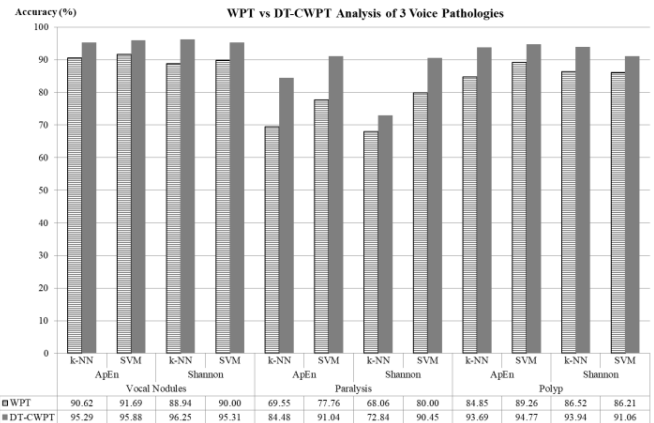


Figure 4 Performances of the WPT and DT-CWPT analysis of 3 voice pathologies after GA

DT-CWPT after GA also shows a better accuracy within classes than WPT as summarizes in Figure 4 for the respective entropies and classifiers. The features produced from the proposed method are more detail and relevant compared to WPT since the algorithm used both real and imaginary tree. In comparison to k-NN, the SVM classifier has achieved better performance of 95.88 %, 91.04% and 94.77% for vocal nodules, paralysis and polyp accordingly. SVM provides a good out-of-sample generalization with proper parameters to achieve a flexible threshold in separating the classes. It can produce a unique solution, unlike neural networks as also discovered by Muhammad [28].

Amongst all the analysis, DT-CWPT achieved the best result with the selected number of features by GA. Performance amongst each class and misclassification seems to distribute evenly using the combination of these feature methods. As mention by Bayram and Selesnick [14], the best basis for the DT-CWPT are more stable compared to the best basis for the WPT when the input signal is shifted. GA find a set of feature subsets that minimize the misclassification rate [29] and provide significant advantages in generalization, at the expense of a

considerable computational cost and a certain bias to the classifier that is used [30]. GA use the chosen learning algorithm or the classifier performance feedbacks for each candidate feature subset evaluation which has a better fit for the chosen learning algorithm or the classifier. These promising results show the reliability of feature extraction and selection method combination in multiclass voice pathology analysis.

4.0 CONCLUSION

Feature extraction using the proposed DT-CWPT based entropies with GA seems to have a good potential in finding optimized features for multiclass voice pathology analysis. It is observed that the best achieved overall accuracy obtained from DT-CWPT with ApEn entropy, 91.15 % for k-NN and 93.90 % for SVM classifier. Both results show improvement in accuracy obtained after GA algorithms been applied. The feature selection techniques select the optimum features in DT-CWPT to achieve better performances. This new implementation of feature extraction using DT-CWPT could lead to better classification rates aimed at identifying the particular type of voice pathology. In future, more than one database can be used as a possible way to prove the reliability of the proposed wavelet feature extraction with feature selection and classification methods.

Acknowledgement

This work was done with part of commercial data from Massachusetts Eye and Ear Infirmary (MEEI) voice disorders database. The authors fully acknowledged the anonymous reviewers for their valuable comments to make this work viable and effective.

References

- [1] J. C. Saldanha, T. Ananthakrishna, and R. Pinto. 2013. Vocal Fold Pathology Assessment using PCA and LDA. *International Conference on Intelligent System and Signal Processing, ISSP 2013*. 140-144. DOI: <https://doi.org/10.1109/ISSP.2013.6526890>.
- [2] P. Harar, Z. Galaz, J. B. Alonso-Hernandez, J. Mekyska, R. Burget, and Z. Smekal. 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. DOI: <https://doi.org/10.1007/s00521-018-3464-7>.
- [3] N. A. S. Selvakumari and V. Radha. 2017. Voice Pathology Identification: A Survey on Voice Disorder. *I.J. Engineering and Manufacturing*. 2: 39-49. DOI: <https://doi.org/10.5815/ijem.2017.02.04>.
- [4] J. Mekyska, E. Janousova, P. Gomez-Vilda, Z. Smekal, I. Rektorova, I. Eliasova, M. Kostalova, M. Mrackova, J. B. Alonso-Hernandez, M. Faundez-Zanuy and K. López-de-Ipiña. 2015. Robust and Complex Approach of Pathological Speech Signal Analysis. *Neurocomputing*. 167: 94-111. DOI: <https://doi.org/10.1016/j.neucom.2015.02.085>.
- [5] M. Hariharan, K. Polat, and S. Yaacob. 2014. A New Feature Constituting Approach to Detection of Vocal Fold Pathology. *International Journal of Systems Science*. 45(8): 1622-1634. DOI: <https://doi.org/10.1080/00207721.2013.794905>.
- [6] A. Sović and D. Seršić. 2012. Signal Decomposition Methods for Reducing Drawbacks of the DWT. *Engineering Review*. 32(2): 70-77.
- [7] P. Saidi and F. Almasgani. 2015. Voice Disorder Signal Classification Using M-Band Wavelets and Support Vector Machine. *Circuits, Systems and Signal Processing*. 34(8): 2727-2738.
- [8] V. Majidnezhad. 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*. 1-11. DOI: <https://doi.org/10.1186/s13636-014-0046-1>.
- [9] A. Akbari and M. K. Arjmandi. 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. DOI: <https://doi.org/10.1016/j.bspc.2013.11.002>.
- [10] F. N. Che Kassim, H. Muthusamy, V. Vijejan, Z. Abdullah, and R. Abdullah. 2020. Dual-Tree Complex Wavelet Packet Transform for Voice Pathology Analysis. *Pertanika Journal of Science and Technology*. 28(3): 839-858.
- [11] F. N. Che Kassim, V. Vijejan, H. Muthusamy, R. Abdullah, and Z. Abdullah. 2019. Voice Pathology Analysis using DT-CWPT and Relief Algorithm. *Journal of Physics: Conference Series*. DOI: <https://doi.org/10.1088/1742-6596/1372/1/012029>.
- [12] S. Hegde, S. Shetty, S. Rai, and T. Dodderi. 2018. A Survey on Machine Learning Approaches for Automatic Detection of Voice Disorders. *Journal of Voice*. 33(6): 947-e11. DOI: <https://doi.org/10.1016/j.jvoice.2018.07.014>.
- [13] G. Roffo. 2016. Feature Selection Library (MATLAB Toolbox). Retrieved from <https://arxiv.org/abs/1607.01327>.
- [14] I. Bayram and I. W. Selesnick. 2008. On the Dual-Tree Complex Wavelet Packet and M-Band Transforms. *IEEE Transactions on Signal Processing*. 56(6): 2298-2310. DOI: <https://doi.org/10.1109/TSP.2007.916129>.
- [15] P. Chattopadhyay and P. Konar. 2014. Feature Extraction using Wavelet Transform for Multi-class Fault Detection of Induction Motor. *Journal of The Institution of Engineers (India): Series B*. 95(1): 73-81. DOI: <https://doi.org/10.1007/s40031-014-0076-1>.
- [16] A. Al-nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, T.A. Mesallam, M. Farahat, K.H. Malki and M.A. Bencherif. 2017. An Investigation of Multidimensional Voice Program Parameters in Three Different Databases for Voice Pathology Detection and Classification. *Journal of Voice*. 31(1): 113-e9. DOI: <https://doi.org/10.1016/j.jvoice.2016.03.019>.
- [17] H. A. Patil. 2019. Combining Evidences From Variable Teager Energy Source and Mel Cepstral Features for Classification of Normal vs. Pathological Voices. *27th European Signal Processing Conference (EUSIPCO)*. 1-5. DOI: <https://doi.org/10.23919/EUSIPCO.2019.8903126>.
- [18] S. M. Pincus, I. M. Gladstone, and R. A. Ehrenkranz. 1991. A Regularity Statistic for Medical Data Analysis. *Journal of Clinical Monitoring*. 7(4): 335-345. DOI: <https://doi.org/10.1007/BF01619355>.
- [19] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore. 2018. Relief-based Feature Selection: Introduction and Review. *Journal of Biomedical Informatics*. 85: 189-203. DOI: <https://doi.org/10.1016/j.jbi.2018.07.014>.
- [20] J. H. Holland. 1975. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor. 100.
- [21] S. Firdos and K. Umarani. 2016. Disordered Voice Classification using SVM and Feature Selection using GA. *Second International Conference on Cognitive Computing and Information Processing (CCIP)*. 1-6. IEEE. DOI: <https://doi.org/10.1109/CCIP.2016.7802868>.
- [22] B. Oluleye, A. Leisa, J. Leng, and D. Dean. 2014. A Genetic Algorithm-based Feature Selection. *International Journal of Electronics Communication and Computer Engineering*. 5(4): 899-905.
- [23] A. Fethi and F. Mohamed. 2013. Voice Pathologies

- Classification Using GMM And SVM Classifiers. *Recent Advances in Biology, Medical Physics, Medical Chemistry, Biochemistry and Biomedical Engineering*. 65-69.
- [24] M. Al Mojaly, G. Muhammad, and M. Alsulaiman. 2014. Detection and Classification of Voice Pathology using Feature Selection. *IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA)*. 571-577. IEEE. DOI: <https://doi.org/10.1109/AICCSA.2014.7073250>.
- [25] I. Hammami, L. Salhi, and S. Labidi. 2020. Voice Pathologies Classification and Detection Using EMD-DWT Analysis Based on Higher Order Statistic Features, *IRBM*, 1 . 1–11. DOI: <https://doi.org/10.1016/j.irbm.2019.11.004>.
- [26] V. Srinivasan, V. Ramalingam, and V. Sellam. 2012. Classification of Normal and Pathological Voice using GA and SVM. *International Journal of Computer Application*. 60(3): 34-39. DOI: <https://doi.org/10.5120/9675-4102>.
- [27] C. C. Chang and C. J. Lin. 2011. LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*. 2(3): 1-27. DOI: <https://doi.org/10.1145/1961189.1961199>.
- [28] G. Muhammad, Alsulaiman, M., Ali, Z., Mesallam, T. A., Farahat, M., Malki, K. H., Al-nasheri, A. and Bencherif, M. A. 2017. Voice Pathology Detection using Interlaced Derivative Pattern on Glottal Source Excitation. *Biomedical Signal Processing and Control*. 31: 156-164. DOI: <https://doi.org/10.1016/j.bspc.2016.08.002>.
- [29] J. Cai, J. Luo, S. Wang, and S. Yang. 2018. Feature Selection in Machine Learning: A New Perspective. *Neurocomputing*. 300: 70-79. DOI: <https://doi.org/10.1016/j.neucom.2017.11.077>.
- [30] J. A. Gómez-García, L. Moro-Velázquez, and J. I. Godino-Llorente. 2019. On the Design of Automatic Voice Condition Analysis Systems. Part I: Review of Concepts and an Insight to the State of the Art . *Biomedical Signal Processing and Control*. 51: 181-199. DOI: <https://doi.org/10.1016/j.bspc.2018.12.024>.