

PAPER • OPEN ACCESS

Voice Pathology Analysis using DT-CWPT and ReliefF Algorithm

To cite this article: Farah Nazlia Che Kassim *et al* 2019 *J. Phys.: Conf. Ser.* **1372** 012029

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

Voice Pathology Analysis using DT-CWPT and ReliefF Algorithm

Farah Nazlia Che Kassim^{1,3*}, Vikneswaran Vijejan¹, Hariharan Muthusamy²,
Rokiah Abdullah^{1,3} and Zulkapli Abdullah³

¹ School of Mechatronic Engineering, Universiti Malaysia Perlis, 02600 Arau, Perlis

² Department of Electronic Engineering, National Institute of Technology, Srinagar (Garhwal), Uttarakhand India

³ Pusat Kejuruteraan, Universiti Malaysia Perlis, 02600 Arau, Perlis

*nazlia@unimap.edu.my

Abstract. Voice pathology analysis has been one of the useful tools in the diagnosis of the pathological voice. This method is non-invasive, inexpensive and reduces time required for analysis. This paper investigates the feature extraction based on the Dual-Tree Complex Wavelet Packet Transform (DT-CWPT) with entropies and energy measures tested with two classifiers, k-Nearest Neighbors (k-NN) and Support Vector Machine (SVM). Feature selection using ReliefF algorithm is applied to reduce redundancy features set and obtain the optimum features for classification. Massachusetts Eye and Ear Infirmary (MEEI) voice disorders database and Saarbruecken Voice Database (SVD) are used. This research was done on multiclass and by specific pathology. The experimental results automates the process of voice analysis hence produce promising results of the presence of diseases in vocal folds.

1. Introduction

Voice is the sound produced by humans using the lungs and the vocal folds in the larynx, or voice box. The presence of pathologies or diseases in vocal folds can affect the normal vibratory patterns, which will affect the sound produced. Some of the voice pathologies are like dysphonia, laryngeal paralysis, cancer, nodules, vocal polyps, vocal fold paralysis, leukoplakia (keratosis), ulcers and voice pathologies due to neurodegenerative disorders like Parkinson disease. A doctor can use the available apparatus such as electromyography (EMG), electroglottography (EGG) and video endoscopy to assess and diagnose malfunctioning vocal fold. However, these methods are known as invasive, high cost and require an expert to analyze the human speech signal parameters. Therefore, a significant attention has been paid to find various parameters obtain from voice signal hence produce early detection results of the presence of diseases in vocal folds, low cost and non-invasive diagnostic technique.

Many automatic systems for the voice pathologies detection have been developed based on voice signal processing. Nowadays, feature extraction using wavelet packet analysis seems to have the reliable capability for the identification of vocal fold pathology [1-5]. The wavelet was preferred because it can provide accurate information about the fast fluctuations of the signals in the time domain [1]. Wavelet packet transform (WPT) is chosen as speech signal parameterization method



because of its ability to analyze a signal at several levels of chosen resolution, considering simultaneously overall information of the spectral band for the speech signal which provides more features and characteristics of the time-frequency characteristics [2].

A typical automatic voice analysis system [4] consist extraction of relevant parameters, decision machines and performance assessment. Feature selection becomes an essential part of a successful data mining pipeline. It is important to discard the irrelevant, redundant and to retain the relevant features to improve the performance of a learning model. At the same time, the time and memory consumed to train a learning model are also reduced. Three major approaches of feature selection methods are filter methods, wrapper methods and embedded methods. The filter methods is renowned to reduced computational burden compared to both embedded and wrapper techniques [4]. Apart from that this methods are independence of the classifier, fast and has good generalization ability. Among the most used filter-based strategies, ReliefF is an iterative, randomized, and supervised approach that estimates the quality of the features according to how well their values differentiate data samples that are near to each other [6]. Other feature selection methods that commonly used in dimensionality reduction technique are fisher method, Laplacian score, genetic algorithms and etc.

In most researches, a two-class classification is done to identify healthy or pathological voice on various databases. Classification is carried out using techniques like vector quantization, artificial neural networks, SVM, hidden Markov model, Gaussian Mixture Model, decision trees, linear classifier, K-means clustering and combined classifiers. Multiclass analysis is less explored due to limited number of available databases with different set of speech tasks and vowel pronounced, labeled in different languages, gender or age and a few type of voice pathologies sample was unevenly distributed across the databases. It is difficult to compare their findings, since their results vary between published papers mostly because of differences in selected voice pathology samples, acoustic features employed, databases and classifiers involved in the research. Selvakumari and Radha [3] concluded in their comparative study between seven researches work mentioned that the acoustic features play the main role in finding the pathology. Therefore, this paper aims are to investigate the use of DT-CWPT for analyzing the voice signals using energy and entropy measures, implementing ReliefF algorithm to select relevant features, performed multiclass classification hence produce results of the presence of diseases in vocal folds.

2. Database

The voice signals from normal person and patients suffering from disorders were acquired from recordings. The voice samples were taken from two databases, MEEI and SVD database for the input signal processing. MEEI database, the most widely used and the only one commercially available become as a benchmark database in the field of pathological speech analysis [7]. The SVD database, a freely downloadable database was recorded by the Institute of Phonetics of Saarland University [8]. Only few studies of voice pathology analysis have been explored on this database [9,10]. Two databases were used as to compare the dependency of voice signal between different language, English and Germany. Table 1 show the number of voice samples uses in the investigations of multiclass analysis for the two different databases. These pathological samples were chosen as they were available and common between the two databases.

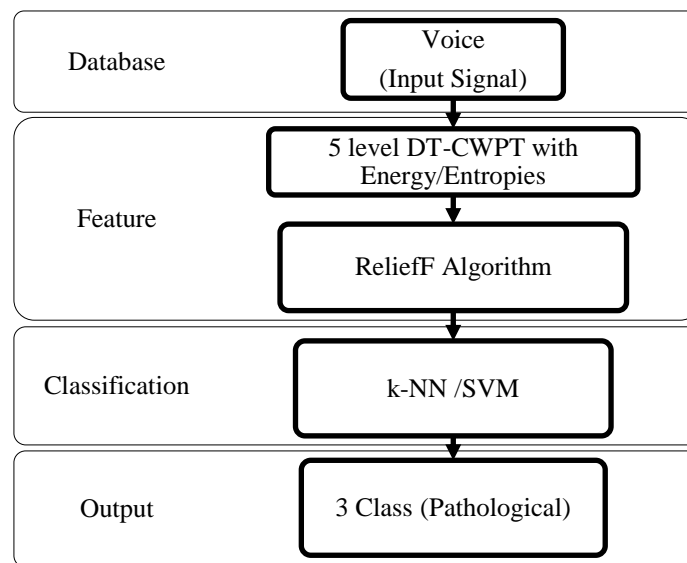
The voice signals files contain sustained normal pitch vowel /a/ samples were downsampled to 25 kHz frequency sampling. An adaptive synthetic (ADASYN) sampling approach is applied to these sets of imbalanced class data to balance up the minority sample to achieve better accuracy. The ADASYN approach improves learning by reducing the bias introduced by the class imbalance and adaptively shifting the classification decision boundary toward the difficult examples [11]. After ADASYN, samples for each class have been balance up to give a total of 201 and 592 samples for MEEI and SVD database respectively.

Table 1. Number of voice samples for multiclass pathology analysis database.

Pathology Class	MEEI Database		SVD Database	
1	Vocal nodules	19	Cysts	6
2	Paralysis	67	Paralysis	194
3	Polyp	20	Polyp	44
Total Samples		106	244	

3. Methodology

The block diagram of the voice pathology analysis is shown in figure 1 consisting DT-CWPT with energy or entropies as feature extraction, feature selection using ReliefF algorithm and multiclass classification by using k-NN and SVM classifier.

**Figure 1.** Block diagram of the voice pathology multiclass analysis.

3.1. Dual-tree Complex Wavelet Packet Transform

The DT-CWPT using energy and entropy measures are investigated as the feature extraction. The DT-CWPT provides time-frequency analysis of the voice signals. It is an extended algorithm from Dual Tree Complex Wavelet Transform (DT-CWT), with two band of DWPT operating in parallel [18]. DT-CWPT has the same shift-invariance and good directional selectivity properties of DT-CWT, and it also has fewer energy leakages into its negative frequency bands [12,13]. Each voice signal is decomposed using the fifth level DT-CWPT and further calculated with energy and non-linear entropy measures, namely Shannon and Renyi entropy which produced a total of 64 ($2^5 \times 2$) features.

3.2. ReliefF Algorithm

This algorithm is implemented after feature extraction. The selection of features was based on the ranking and weightage computed by ReliefF algorithm which list the most useful to least useful features. The original relief can deal with nominal and numerical attributes. However, it cannot deal with incomplete data and is limited to two-class problems. ReliefF notably, the 'F' refers to the sixth algorithm variation (from A to F) proposed by Kononenko [14] is extended to respond to deal with multiclass problems, incomplete and noisy data and are more robust. Instead of finding one near miss M from different class, the algorithm finds one near miss M(C) for each different class and averages their contribution for updating estimates W[A]. The average is weighted with the prior probability of each class as in equation (1).

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

For $i=1$ to m times, ReliefF randomly selects an instance R_i , find k nearest hits H_j and misses $M_j(C)$. The k is tuned to 10 (optimal k for dataset) [14-16]. The values of k is also varies to investigate the stability and reliability of attribute ranks and weights. It updates the quality estimation $W[A]$ for all attributes A depending on their values for R_i , hits H_j and misses $M_j(C)$. The contribution for each class of the misses is weighted with the prior probability of that class $P(C)$.

3.3. Classification

Two common classifiers, k -NN and SVM were used. k -NN is one of the simplest machine learning algorithms, also called as lazy learner [5]. Classification was based on majority of k -Nearest Neighbor's category and use distance measure to decide in the train data. k values were varied between 1 and 10 for a given training and test data. A 10-fold cross validation classification (CVC) scheme was used to increase the reliability of the results [17,18]. The SVM with radial basis function (RBF) as a kernel is used. This classifier have several advantages such as it can provide a good out-of-sample generalization, it can produce a unique solution if the parameters are properly chosen, unlike neural networks, because the optimality problem is convex, and with the help of kernels, it can achieve the flexibility of the form of the threshold separating the classes [11].

Various experiments had been made, aiming to select the best number of features for classification as well as reducing the time taken for classification. First experiment is the classification using all the features from DT-CWPT with energy features, Shannon entropy and Renyi entropy features which produced 64 features each. Next experiment involved reducing the features by selecting top most relevant 10, 20 and 30 features as until almost half of the original features. The accuracy and time taken for each classifier were then observed and compared.

4. Results and Discussion

Table 2 and Table 3 summarises the multiclass results obtained from the experiments for both database. Table 2 shows the accuracy and classification time taken to classify three pathological class MEEI database using DT-CWPT with energy, Shannon and Renyi entropy. Using k -NN classifier, the highest prediction rates were obtained by selecting the top 30 features. The best achieved accuracy for k -NN is 80%, 80.10% and 79.55% for DTCWPT with energy, Shannon and Renyi entropy respectively. These accuracy increases 1-6% compared to using all the 64 features with the advantages of reduced the classification time taken. In contrast using SVM classifier, the accuracy reduced about 2.5% for DT-CWPT with Shannon entropy for chosen 30 features. The prediction rate of DT-CWPT with energy and Renyi entropy increase and are among the highest achieved, 93.63% and 94.03%. ReliefF give the advantage of reducing almost 30% of the classification time.

The classification results for SVD database using DT-CWPT with energy, Shannon and Renyi entropy were shown in table 3. Better accuracy achieved, 85.81- 87.03% about 3% increase using selected 30 features for k -NN classifier. Average calculation time reduce about 3-11% compared to use all features. In contrast to the SVM classifier results, the highest prediction rate is 97.65% by taking all features from DT-CWPT with Renyi entropy but takes a longer time. A slight 1% reduces of accuracy achieved for the selected 30 features but time for calculation become faster.

Comparing both of the classifiers, SVMs give higher accuracy than k -NN. SVM classifier is used most widely due to the fact that SVM classifier performs better for high-dimensional data as well as small sized dataset [5] but SVMs intensive computation takes a longer time to make decision. Both class 2 in k -NN classifier show a low rate maybe due to accuracy of regions declines for higher dimensional data sets. It determines the class by taking the majority vote of class labels among the k -nearest neighbors (uses all features in computing distances).

ADASYN was used to overcome the small samples, this give influence to class 1 SVD database which gives perfect prediction rate. In future studies, the findings from synthetic data could be validated with real samples. Overall, the performances of both classifiers using the dataset after feature selection are better and improved in term of both accuracy and time taken.

Table 2. Comparison of accuracy and classification time for three class MEEI database.

Classifier	Features Experiments	Accuracy (%) \pm sd			Time (seconds)		
		Average	Class 1	Class 2		Class 3	
KNN	DT-CWPT + Energy						
	30 features	80.00 \pm 0.77	97.19 \pm 0.66	46.57 \pm 1.83	96.29 \pm 1.54	1.61	
	64 features	79.25 \pm 0.47	97.19 \pm 0.99	44.78 \pm 1.00	95.86 \pm 0.45	1.62	
	DT-CWPT + Shannon entropy						
	30 features	80.10 \pm 0.81	96.88 \pm 0.00	46.12 \pm 2.38	97.29 \pm 1.05	0.63	
	64 features	75.82 \pm 1.05	96.56 \pm 1.23	34.03 \pm 2.88	96.86 \pm 0.60	0.64	
	DT-CWPT + Renyi entropy						
	30 features	79.55 \pm 0.95	96.72 \pm 1.15	45.22 \pm 1.42	96.71 \pm 0.96	0.62	
	64 features	79.10 \pm 0.91	97.03 \pm 0.89	44.48 \pm 2.97	95.86 \pm 0.81	0.63	
	SVM	DT-CWPT + Energy					
		30 features	93.63 \pm 1.04	92.03 \pm 2.70	94.03 \pm 0.70	94.71 \pm 1.18	13.81
		64 features	93.48 \pm 0.92	90.63 \pm 2.33	97.01 \pm 0.00	92.71 \pm 1.25	19.37
DT-CWPT + Shannon entropy							
30 features		89.90 \pm 1.00	91.88 \pm 1.77	88.66 \pm 1.04	89.29 \pm 2.26	13.66	
64 features		92.19 \pm 0.78	89.38 \pm 1.92	98.36 \pm 0.85	88.86 \pm 1.31	19.44	
DT-CWPT + Renyi entropy							
30 features		94.03 \pm 0.66	93.44 \pm 1.23	94.03 \pm 1.00	94.57 \pm 1.88	13.60	
64 features		93.53 \pm 0.70	90.47 \pm 2.01	97.16 \pm 0.47	92.86 \pm 1.51	19.20	

Table 3. Comparison of accuracy and classification time for three class SVD database.

Classifier	Features Experiments	Accuracy (%) \pm sd			Time (seconds)		
		Average	Class 1	Class 2		Class 3	
KNN	DT-CWPT + Energy						
	30 features	86.42 \pm 0.36	100.00 \pm 0.00	62.58 \pm 0.92	96.18 \pm 0.39	1.64	
	64 features	83.75 \pm 0.40	100.00 \pm 0.00	52.58 \pm 0.97	97.94 \pm 0.45	1.71	
	DT-CWPT + Shannon entropy						
	30 features	85.81 \pm 0.51	100.00 \pm 0.00	59.74 \pm 1.25	97.11 \pm 0.59	0.63	
	64 features	83.85 \pm 0.44	100.00 \pm 0.00	53.40 \pm 1.01	97.45 \pm 0.56	0.72	
	DT-CWPT + Renyi entropy						
	30 features	87.03 \pm 0.54	100.00 \pm 0.00	63.35 \pm 1.69	97.21 \pm 0.61	0.70	
	64 features	84.34 \pm 0.35	100.00 \pm 0.00	54.43 \pm 1.24	97.89 \pm 0.33	0.78	
	SVM	DT-CWPT + Energy					
		30 features	96.55 \pm 0.35	100.00 \pm 0.00	97.53 \pm 0.33	92.35 \pm 1.23	98.00
		64 features	97.45 \pm 0.26	100.00 \pm 0.00	98.97 \pm 0.00	93.58 \pm 0.75	148.31
DT-CWPT + Shannon entropy							
30 features		93.77 \pm 0.30	100.00 \pm 0.00	85.05 \pm 0.81	96.13 \pm 0.82	98.53	
64 features		94.81 \pm 0.28	100.00 \pm 0.00	86.60 \pm 0.49	97.70 \pm 0.40	147.13	
DT-CWPT + Renyi entropy							
30 features		96.55 \pm 0.35	100.00 \pm 0.00	97.53 \pm 0.33	92.35 \pm 1.23	97.48	
64 features		97.69 \pm 0.33	100.00 \pm 0.00	96.60 \pm 0.36	96.52 \pm 0.71	149.96	

5. Conclusion

In conclusion, ReliefF algorithm can be an effective way to improve the overall performance of the multiclass classification. It can enhance the DT-CWPT performance by selecting relevant features and reduce time taken for accuracy calculation. For future works, the system performance should be tested

by increasing the training samples and the used of another database to assess the independence of the developed algorithms. From the promising experimental results, it is hoped that the system can be test rigorously in the medical field and provide the clinicians with a useful early diagnosis for voice pathology.

References

- [1] Taib,D., Tarique, M., Islam, R.: Voice feature analysis for early detection of voice disability in children. *IEEE International Symposium on Signal Processing and Information Technology* (2018) 12-17
- [2] Arjmandi, M.K., Pooyan, M.: An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine. *Biomedical Signal Processing and Control* 7(1) (2012) 3–19
- [3] Selvakumari, N.A.S., Radha, V.: Voice pathology identification: a survey on voice disorder. *International Journal of Engineering and Manufacturing* 7(2) (2017) 39–49.
- [4] Gómez-garcía, J.A., Moro-velázquez, L., Godino-llorente, J.I.: 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 (2019) 181–199
- [5] Hegde, S., Shetty, S., Rai, S., Dodderi, T.: A survey on machine learning approaches for automatic detection of voice disorders. *Journal of Voice* (2018)
- [6] Roffo, G., Melzi, S., Castellani, U., Vinciarelli, A.: Infinite latent feature selection: a probabilistic latent graph-based ranking approach. In: *Proceedings of the IEEE International Conference on Computer Vision* (2017) 1398-1406
- [7] Harar, P., Galaz, Z., Alonso-Hernandez, J.B., Mekyska, J., Burget, R., Smekal, Z.: Towards robust voice pathology detection. *Neural Computing and Applications* (2018) 1–11
- [8] Barry, W.J., Pützer, M.: Saarbruecken voice database. Institute of Phonetics, University of Saarland, <http://www.stimmdatenbank.coli.uni-saarland.de>. (2007)
- [9] Martinez, D., Lleida, E., Ortega, A., Miguel, A., Villalba, J.: 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* Springer, Berlin, Heidelberg (2012) 99–109
- [10] Muhammad, G., Alsulaiman, M., Ali, Z., Mesallam, T.A., Farahat, M., Malki, K.H., Bencherif, M.A.: Voice pathology detection using interlaced derivative pattern on glottal source excitation. *Biomedical Signal Processing and Control* (2017) 156–164.
- [11] He, H., Bai, Y., Garcia, E.A, Li, S.:ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In *IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)* (2008) 1322–1328
- [12] Bayram, I., Selesnick, I.W.: On the dual-tree complex wavelet packet and m-band transforms. *IEEE Transactions on Signal Processing* 56(6) (2008) 2298–2310
- [13] Selesnick, I.W., Baraniuk, R.G., Kingsbury, N.G.: The dual-tree complex wavelet transform. *IEEE signal processing magazine* 22(6) (2005) 123-151.
- [14] Kononenko, I.: Estimating attributes: analysis and extensions of relief. *European Conference on Machine Learning*, Springer, Berlin, Heidelberg (1994) 171–182
- [15] Roffo,G.:Feature selection library (MATLAB toolbox).arXiv preprint arXiv:1607.01327 (2016)
- [16] Robnik-Šikonja, M., Kononenko, I.: Theoretical and empirical analysis of ReliefF and RReliefF. *Machine learning* 53 (1-2) (2003) 23-69
- [17] Abdullah, R., Muthusamy, H., Vijean, V., Abdullah, Z., Kassim, F.N.C.: Real and complex wavelet transform approaches for malaysian speaker and accent recognition. *Pertanika Journal of Science & Technology* 27(2) (2019) 737–752
- [18] Hariharan, M., Paulraj, M.P., Yaacob, S.:Time-domain features and probabilistic neural network for the detection of vocal fold pathology. *Malaysian Journal of Computer Science* 23(1) (2010) 60–67