



**Implementation of Intelligent Signal Processing
Method for Vocal Fold Pathology Detection through
Speech Signals**

by

**Farah Nazlia Che Kassim
(1431311444)**

A thesis submitted in fulfillment of the requirements for the degree of
Master of Science in Biomedical Electronic Engineering

**Faculty of Electronic Engineering Technology
UNIVERSITI MALAYSIA PERLIS**

2021

ACKNOWLEDGMENT

Alhamdulillah, all praises to Allah s.w.t, Almighty God I managed to complete my master thesis after all the hardship and tribulations during this adventurous journey with His blessing and guidance. I would like to take this opportunity to acknowledge and appreciate the efforts of all people involved that helped me during my study and the documentation of this thesis. Special thanks to my supervisors, especially Dr. Vikneswaran Vijejan, Dr. Hariharan Muthusamy, Prof Sazali Yaacob, Prof Madya Che Mohd Ruzaidi Ghazali, Prof Madya Abdul Rahman, my colleagues at Pusat Kejuruteraan UniMAP especially Mr. Zulkapli Abdullah and Mrs. Rokiah Abdullah, the Centre for Graduate Studies, Faculty of Electronic Engineering Technology UniMAP Deans and staffs, my parents (Mr. Che Kassim Hasan and Mrs. Siti Hawa Ismail), family and friends and those that I overlooked not mentioning who always share experiences and knowledge, keep me motivated and contribute, a huge thanks.

May Allah grant them the best rank in this world and hereafter.

TABLE OF CONTENTS

	PAGE
DECLARATION OF THESIS	i
TABLE OF CONTENTS	iii
LIST OF TABLES	vii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	xii
LIST OF SYMBOLS	xiv
ABSTRAK	xv
ABSTRACT	xvi
CHAPTER 1 : INTRODUCTION	1
1.1 Introduction	1
1.2 Problem Statement	2
1.3 Research Objectives	4
1.4 Scope of Research	4
1.5 Thesis Outline	5
CHAPTER 2 : LITERATURE REVIEW	6
2.1 Introduction	6
2.2 Vocal Fold Pathology	6
2.3 Databases	8
2.4 Previous Works	9
2.5 Feature Extraction	13

2.5.1	Wavelet	13
2.5.2	DT-CWPT	16
2.6	Feature Selection	18
2.7	Classification	19
2.8	Research Gap	22
CHAPTER 3 : METHODOLOGY		24
3.1	Introduction	24
3.2	Database	27
3.2.1	MEEI Database	27
3.2.2	SVD Database	27
3.2.3	Experimental Dataset	28
3.2.4	Voice Signal Pre-Processing	30
3.3	Feature Extraction	32
3.3.1	Mel-Frequency Cepstral Coefficients	33
3.3.2	Linear Predictive Coding	34
3.3.3	Wavelet Packet Transform	35
3.3.4	Dual-tree Complex Wavelet Packet Transform	37
3.3.5	Wavelet Decomposition Level	39
3.3.6	Energy and Non-linear Entropy Features	40
3.3.7	ADASYN	43
3.4	Feature Selection	44
3.4.1	ReliefF Algorithm	45
3.4.2	Genetic Algorithm	48
3.5	Classification	49
3.5.1	<i>k</i> -NN Classifier	49
3.5.2	SVM Classifier	50
3.5.3	Cross Validation	52

3.5.4	Pairwise Analysis	52
3.5.5	Multiclass Analysis	54
CHAPTER 4 :	RESULT AND DISCUSSION	55
4.1	Overview	55
4.2	Pairwise Analysis Results	56
4.2.1	Comparison with Previous Work	63
4.3	Multiclass Analysis Results	64
4.3.1	Dataset 4	65
4.3.1.1	File-based vs Frame-based before Feature Selection	70
4.3.1.2	File-based vs Frame-based after Feature Selection	72
4.3.1.3	Accuracies within Classes	74
4.3.2	Dataset 5	77
4.3.2.1	File-based vs Frame-based before Feature Selection	82
4.3.2.2	File-based vs Frame-based after Feature Selection	84
4.3.2.3	Accuracies within Classes	86
4.3.3	Comparison with Previous Work	93
4.4	Results Summary	94
CHAPTER 5 :	CONCLUSION	97
5.1	Summary	97
5.2	Future Work	99
REFERENCES		100
APPENDIX A	Level Differences of Wavelet	111
APPENDIX B	Accuracy Before and After ADASYN	112
APPENDIX C	ReliefF and GA Parameters Optimization	113

APPENDIX D Sensitivity and Specificity Of Pairwise Analysis	114
LIST OF AWARDS AND PUBLICATIONS	115

©This item is protected by original copyright

LIST OF TABLES

	PAGE	
Table 2.1	Description of some vocal fold pathologies.	7
Table 2.2	Voice pathology database.	8
Table 2.3	Summary of previous work in vocal fold pathology detection.	10
Table 2.4	Significant works using wavelets.	14
Table 2.5	Significant DT-CWPT works.	16
Table 2.6	Significant multiclass work.	19
Table 3.1	Numbers of voice samples for pairwise analysis and multiclass analysis.	29
Table 3.2	Number of voice samples after ADASYN	44
Table 3.3	Parameters used in GA	49
Table 3.4	Confusion matrix.	53
Table 3.5	Definition of the terminology.	53
Table 3.6	Definition of measures of performance.	53
Table 3.7	Multiclass confusion matrix.	54
Table 4.1	Results overview.	55
Table 4.2	Result of k -NN classifier for pairwise analysis.	57
Table 4.3	Result of SVM classifier for pairwise analysis.	59
Table 4.4	Overview of pairwise analysis using MEEI database (53 Normal and 173 Pathological).	63

Table 4.5	Result of k -NN classifier for Dataset 4 (File-based).	66
Table 4.6	Result of k -NN classifier for Dataset 4 (Frame-based).	67
Table 4.7	Result of SVM classifier for Dataset 4 (File-based).	68
Table 4.8	Result of SVM classifier for Dataset 4 (Frame-based).	69
Table 4.9	Result of k -NN classifier for Dataset 5 (File-based).	78
Table 4.10	Result of k -NN classifier for Dataset 5 (Frame-based).	79
Table 4.11	Result of SVM classifier for Dataset 5 (File-based).	80
Table 4.12	Result of SVM classifier for Dataset 5 (Frame-based).	81
Table 4.13	Results summary of accuracy and number of features after GA.	92
Table 4.14	Accuracy of the methods for multiclass analysis (File-Based).	93
Table 4.15	Results summary of DT-CWPT.	94

LIST OF FIGURES

	PAGE
Figure 1.1 Human voice production system (Rosistem, 2014)	1
Figure 1.2 Laryngoscopy (Brady & Becker, 2016)	2
Figure 2.1 The general block diagram for vocal fold pathology detection.	6
Figure 3.1 Overall methodology for vocal fold pathology detection.	24
Figure 3.2 Block diagram of the pairwise vocal fold pathology detection.	25
Figure 3.3 Block diagram of the multiclass vocal fold pathology detection.	26
Figure 3.4 Data distribution for each experimental dataset.	28
Figure 3.5 Framing and overlapping frame arrangement of a sample voice signal.	31
Figure 3.6 Samples of frame-based waveform from a voice signal.	32
Figure 3.7 Block diagram of the MFCC (Alim & Rashid,2018)	33
Figure 3.8 Block diagram of the LPC (Alim & Rashid,2018)	34
Figure 3.9 WPT for three levels (Burrus et al., 1998)	36
Figure 3.10 First wavelet packet FB of 5-level DT-CWPT (Bayram & Selesnick, 2008).	38
Figure 3.11 Rank for features derived from DT-CWPT with energy.	47

Figure 3.12	Rank for features derived from DT-CWPT with Tsallis entropy.	47
Figure 3.13	k -NN algorithms (Goel & Mahajan, 2017)	50
Figure 3.14	The binary (pairwise) and multiclass classification cases (Polat & Nour, 2020)	51
Figure 4.1	Summary of pairwise analysis using k -NN classifier.	58
Figure 4.2	Summary of pairwise analysis using SVM classifier.	60
Figure 4.3	Sensitivity of the pairwise analysis.	62
Figure 4.4	Specificity of the pairwise analysis.	63
Figure 4.5	Summary of file-based vs frame-based before feature selection for Dataset 4.	71
Figure 4.6	Average multiclass accuracy for Dataset 4 (File-based).	72
Figure 4.7	Average multiclass accuracy for Dataset 4 (Frame-based).	73
Figure 4.8	Comparison within classes with GA for Dataset 4 (File-based).	74
Figure 4.9	Comparison within classes with GA for Dataset 4 (Frame-based).	75
Figure 4.10	Comparison within classes with ReliefF for Dataset 4 (File-based).	75
Figure 4.11	Comparison within classes with ReliefF for Dataset 4 (Frame-based)	76
Figure 4.12	Summary of file-based vs frame-based before feature selection for Dataset 5.	83
Figure 4.13	Average multiclass accuracy for for Dataset 5 (File-based).	84

Figure 4.14	Average multiclass accuracy for for Dataset 5 (Frame-based).	85
Figure 4.15	Comparison within classes with GA for Dataset 5 (File-based).	86
Figure 4.16	Comparison within classes with GA for Dataset 5 (Frame-based).	87
Figure 4.17	Comparison within classes with ReliefF for Dataset 5 (File-based).	87
Figure 4.18	Comparison within classes with ReliefF for Dataset 5 (Frame-based).	88
Figure 4.19	Features distribution between three classes before GA.	89
Figure 4.20	Features distribution between three classes after GA.	89
Figure 4.21	The comparison of k -NN and SVM classification time analysis.	92

LIST OF ABBREVIATIONS

ADASYN	Adaptive Synthetic
ANN	Artificial Neural Networks
ApEn	Approximate Entropy
AVFAD	Advanced Voice Function Assessment Databases
AVPD	Arabic Voice Pathology Database
CVC	Cross-Validation Classification
db	Daubechies
DCT	Discrete Cosine Transform
DT-CWPT	Dual-Tree Complex Wavelet Packet Transform
DT-CWT	Dual-Tree Complex Wavelet Transform
DWPT	Discrete Wavelet Packet Transform
DWT	Discrete Wavelet Transform
FB	Filter Banks
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FN	False Negative
FP	False Positive
GA	Genetic Algorithm
GMM	Gaussian Mixture Model
GRNN	General Regression Neural Network
HMM	Hidden Markov Model
HOS	Higher Order Statistics
IDP	Interlaced Derivative Pattern
k -NN	k -Nearest Neighbours
LDA	Linear Discriminant Analysis
LPC	Linear Predictive Coding
LPCC	Linear Prediction Cepstral Coefficient
MDVP	Multidimensional voice program analysis
MEEI	Massachusetts Eye and Ear Infirmary
MFCC	Mel Frequency Cepstral Coefficients
MLPNN	Multilayer Perceptron Neural Network
MS	Modulation Spectra
OSAHS	Obstructive Sleep Apnea-Hypopnea Syndrome

PARCZ	Czech Parkinsonian Speech
PCA	Principal Component Analysis
PdA	Principe De Asturias
PNN	Probabilistic Neural Network
PR	Perfect Reconstruction
RAM	Random Access Memory
RBF	Radial Basis Function
RF	Random Forest
SD	Standard Deviation
SGD	Stochastic Gradient Descent
STFT	Short-Time Fourier Transform
SVD	Saarbruecken Voice Database
SVM	Support Vector Machine
SWT	Stationary Wavelet Transform
TN	True Negative
TP	True Positive
TQWT	Tunable Q-factor Wavelet Transform
WPT	Wavelet Packet Transform

©This item is protected by original copyright

LIST OF SYMBOLS

$e(n)$	Residual error
f	Frequency
h_0	Low-pass filter
h_1	High-pass filter
j	Level of decomposition
m	Embedding dimension
N_f	Cardinality of the selected features
p	Order of prediction
r	Filtering level (threshold)
α	Nonextensivity parameter
γ	Gamma (regularisation parameter)
σ^2	Sigma (squared bandwidth of the RBF kernel)

©This item is protected by original copyright

Pelaksanaan Kaedah Pintar Pemprosesan Isyarat untuk Mengesan Patologi Suara melalui Isyarat Pertuturan

ABSTRAK

Manusia menghasilkan suara melalui paru-paru dan peti suara. Kehadiran patologi atau penyakit pada peti suara boleh memberi kesan pada corak getaran normal dan kualiti suara. Pengesanan penyakit peti suara adalah tugas yang sukar kerana beberapa diagnosis semasa seperti *laryngoscopy* dan *video endoscopy* bersifat invasif dan memerlukan kepakaran untuk menganalisis parameter suara manusia. Kaedah diagnostik tidak invasif berdasarkan pemprosesan suara iaitu sistem pengesanan penyakit peti suara secara automatik untuk mencari pelbagai parameter yang diperoleh dari suara adalah dicadangkan. Kaedah ini tidak invasif dan berguna sebagai alat pengesanan untuk membantu doktor membuat keputusan yang lebih baik mengenai keadaan pesakit untuk pemeriksaan klinikal selanjutnya. Sebilangan besar penyelidik menggunakan ciri-ciri dalam domain frekuensi untuk mencari parameter suara. Walaubagaimanapun, terdapat beberapa batasan ditemui seperti kehilangan maklumat domain masa semasa melakukan transformasi frekuensi sehingga sukar untuk menentukan dan mendiagnosis gangguan suara tertentu secara klinikal. Sebilangan besar kajian juga terhad kepada klasifikasi dua kelas dan kurang memberi tumpuan kepada multikelas. Kajian ini menyelidik prestasi ciri-ciri daripada pengekstrakan *Dual-Tree Complex Wavelet Packet Transform* (DT-CWPT) dengan ukuran tenaga dan entropi yang dikuantifikasi dengan dua pengklasifikasi, *k-Nearest Neighbours* (*k*-NN) dan *Support Vector Machine* (SVM). DT-CWPT telah berjaya dilaksanakan dalam pelbagai bidang kerana memperkenalkan ciri-ciri kompleks dalam skala masa dan frekuensi yang memberikan resolusi masa dan frekuensi yang lebih baik, *shift-invariance* dan *good directional selectivity* di mana ciri-ciri ini kurang dalam transformasi wavelet yang tradisional. Penguraian komposisi suara dilakukan dengan menggunakan ukuran tenaga dan entropi untuk mengurangkan kerumitan isyarat suara dalam domain masa dan frekuensi. Lima set data yang diperoleh dari pangkalan data gangguan suara *Massachusetts Eye and Ear Infirmary* (MEEI) dan *Saarbruecken Voice Database* (SVD) digunakan. Pemilihan ciri-ciri menggunakan algoritma *ReliefF* dan *Genetic algorithm* (GA) dicadangkan ke atas set data multikelas untuk mengurangkan redundansi dan mendapatkan ciri optimum untuk klasifikasi. Proses pengesanan patologi suara secara automatik ini dianalisis berasaskan fail dan berasaskan kerangka secara dua kelas (*normal and abnormal class*) dan multikelas yang melibatkan beberapa penyakit berkaitan peti suara (*Vocal Fold Cysts, Vocal Nodules, Polyp, and Paralysis*). Untuk analisis dua kelas, prestasi terbaik yang diperoleh adalah dengan menggunakan kombinasi DT-CWPT dan SVM yang mencapai ketepatan 99.73% sehingga 100% sementara ketepatan 93.90% sehingga 99.64% dicapai untuk multikelas. Kaedah yang dicadangkan ini memberikan kadar pengesanan yang baik untuk mengesan pelbagai penyakit gangguan suara dan berguna untuk pembangunan alat diagnostik berasaskan komputer untuk pemeriksaan patologi suara di kemudahan penjagaan kesihatan.

Implementation of Intelligent Signal Processing Method for Vocal Fold Pathology Detection through Speech Signals

ABSTRACT

Humans produce voice through lungs and vocal folds or vocal cords in the larynx. The presence of pathologies or diseases in vocal folds could have an impact on the normal vibratory patterns and sound quality. The vocal fold pathology detection was a difficult task because some current diagnosis such as laryngoscopy and video endoscopy are invasive and it needs expert and exact knowledge to analyze the human voice signal parameters. A significant non-invasive diagnostic technique based on voice signal processing i.e. an automated vocal fold pathology detection system to find various parameters or features obtained from the voice signal is proposed. This method is non-invasive and can be useful as detection tool to assist doctors to make better decisions on patients' conditions for further clinical examinations. Most researchers used features in frequency-domain to find the voice parameters. A few limitations were discovered such as the loss of time-domain information while performing the frequency transformation making it difficult to define and diagnose specific voice disorders clinically. Also, majority of works are limited to pairwise classification problems and less focus on the classification of the exact pathology. This study investigates the performance of features derived from the Dual-Tree Complex Wavelet Packet Transform (DT-CWPT) with energy and entropies measures quantified with two classifiers, k -Nearest Neighbours (k -NN) and Support Vector Machine (SVM). The DT-CWPT introduces the complex coefficients in time and frequency scale which delivers a better time-frequency resolution, provides approximately shift-invariance and good directional selectivity where these properties are lacking in traditional wavelet transform. Decomposition is done on the voice signals using energy and entropies to signify the complexity of voice signals in the time and frequency domains. Five sets of datasets obtained from the Massachusetts Eye and Ear Infirmary (MEEI) voice disorders database and Saarbruecken Voice Database (SVD) are used. In order to reduce the high dimensionality feature of multiclass datasets, feature selections using ReliefF algorithm and Genetic algorithm (GA) are proposed to reduce redundancy features and obtain the optimum features for classification. This automatic voice pathologies detection experimented with file-based and frame-based analyses for pairwise (normal and abnormal class) and multiclass classification of the abnormal voice pathologies (Vocal Fold Cysts, Vocal Nodules, Polyp, and Paralysis). It is observed that the best accuracies obtained using a combination of DT-CWPT with SVM classifier for the pairwise analysis achieved 99.73% to 100% accuracy meanwhile, 93.90% to 99.64% of average accuracy achieved in multiclass. The presented study 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.

CHAPTER 1 : INTRODUCTION

1.1 Introduction

Humans produce voice through lungs and vocal folds or vocal cords in the larynx as shown in Figure 1.1. The presence of pathologies or diseases in vocal folds could have an impact on the normal vibratory patterns and sound quality. Some example of the voice disorders or pathologies are vocal polyps, ulcers, dysphonia, leukoplakia (keratosis), laryngeal paralysis, nodules, vocal fold paralysis, laryngitis, hyponasality, hypernasality and voice pathologies due to neurodegenerative disorders such as Parkinson disease. Potential peoples suffering from these diseases are those who use voices frequently and prolonged basis, such as singers, teachers, lecturers, lawyers and customer service workers. Smoking habit, intake of drinks containing alcohol or caffeine excessively can also contribute to this health problem. Voice signal analysis is becoming one of the useful alternative diagnosis tools to detect voice abnormalities and assist doctors to make better decisions on patients' conditions for further clinical examinations.

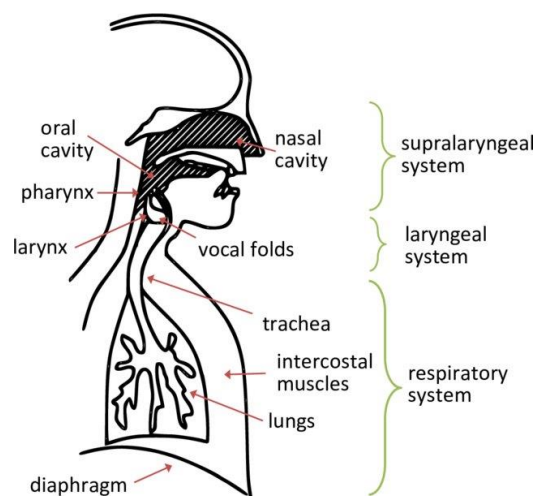


Figure 1.1 Human voice production system (Rosistem, 2014)

1.2 Problem Statement

In the whole group of 349 people from self-assessment questionnaires, the total prevalence of voice disorders was calculated at 16.9%, where 15.5% of voice disorders were rated to occur to a minor extent and 1.4% to a large extent (Lyberg-Åhlander, Rydell, Fredlund, Magnusson, & Wilén, 2019). Voice disorders affect 15% to 30% of the general population of 430 people and have also confirmed poorer quality of life such as limited social activities, more depression, poorer job performance and lower self-esteem (DeVore, Carroll, Rosner, & Shin, 2020). Normally, doctors perform physical examinations and use the available apparatus to assess and diagnose malfunctioning vocal fold by performing procedures such as laryngoscopy as in Figure 1.2, stroboscopy, video endoscopy, electromyography, electroglottography (Stachler et al., 2018). However, some of these methods are invasive, can cause discomfort to patients and require a specialist to examine the parameters of human voice signal (Islam, Tarique, & Abdel-Raheem, 2020; Al-Dhief et al., 2020; Bendale et al., 2021).

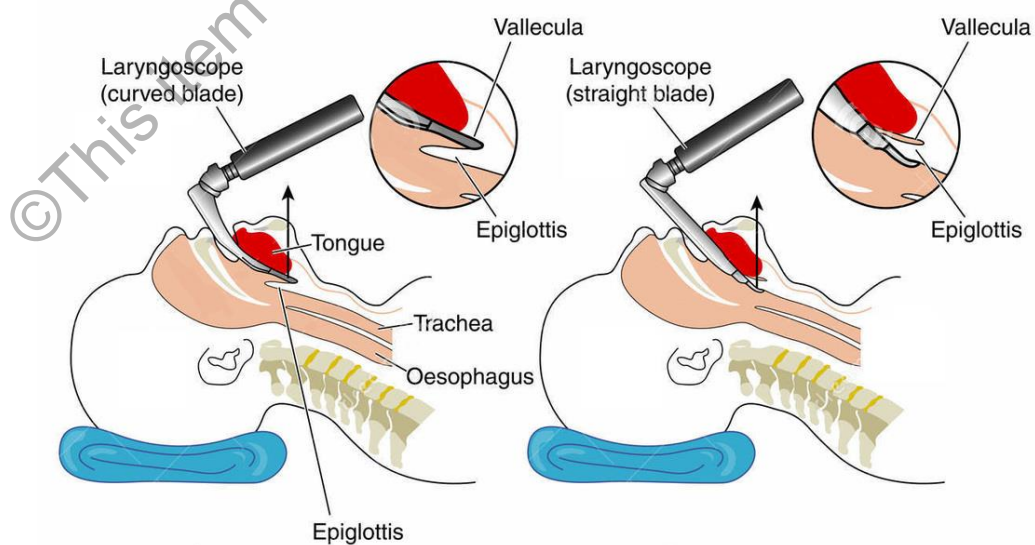


Figure 1.2 Laryngoscopy (Brady & Becker, 2016)

The human voice has been shown to carry much information about the general health and well-being of a person (Titze, 1994). Therefore, a significant non-invasive diagnostic technique based on voice signal processing i.e. an automated vocal fold pathology detection system to find various parameters or features obtained from the voice signal is proposed.

Most of the researchers use the features in the frequency domain though few researchers rely on time-domain measures for a specific pathology such as cysts, polyp, dysphonia, paralysis and vocal nodules (Islam et al.,2020). However features in the frequency domain such as Mel Frequency Cepstral Coefficients (MFCC) and Linear Predictive Coding (LPC) have some limitations present such as notoriously difficult to define clinically and to diagnose specific voice disorders (Mekyska, Janousova, Gomez-Vilda, Smekal, Rektorova, Eliasova, & et al.,2015), low robustness to noise signals (Harar et al., 2018; Alim & Rashid, 2018; Gadekar, Kaldane, Pawar, Jadhav, & Patil, 2019;Bendale et al., 2021) and to identify their early stages or track their progression (Harar et al., 2018). The speech signal is a highly non-stationary signal where the signal spectral content or frequency changes with time. Fourier transform (frequency domain) is ineffective to analyse non-stationary signal since time-domain information is lost during the frequency transformation (Hariharan, Polat, & Yaacob, 2014). The Wavelet Packet Transform (WPT) was discovered by the authors to be preferable for the non-stationary signals investigation both in time and frequency scale.

In the literature, numerous features have been researched using acoustics parameters, features extraction techniques and various machine learning approaches to identify healthy and pathological voices but the challenges are to identify specific

pathology (Islam et al., 2020). Majority of works are limited to pairwise classification problems and less focuses on the classification of the exact the pathology. Based on the literature, this study focused on investigating features derived from wavelet methods namely Dual-Tree Complex Wavelet Packet Transform (DT-CWPT) for pairwise and multiclass analysis in vocal fold pathology detection. The DT-CWPT introduces the complex coefficients in time and frequency scale which delivers a better time-frequency resolution, provides approximately shift-invariance and good directional selectivity where these properties are lacking in traditional wavelet transform.

1.3 Research Objectives

The aims of this study are listed below:

- i To propose a non-invasive method for vocal fold pathology detection through voice signals.
- ii To investigate features derived from Dual-Tree Complex Wavelet Packet Transform (DT-CWPT) and identify suitable feature for the detection of selected vocal fold pathology.
- iii To analyse the performance of the vocal fold pathology detection in pairwise and multiclass analysis.

1.4 Scope of Research

The scope of this study is to investigate the suitability of wavelet-based features in a vocal fold pathology detection system. DT-CWPT is proposed and compared with WPT, MFCC and LPC. Features such as entropies and energy derived from WPT and

DT-CWPT are investigated. Voice signals from selected pathology are taken from two distinguished databases, the commercial Massachusetts Eye and Ear Infirmary (MEEI) voice disorders database and Saarbruecken Voice Database (SVD) database which freely available online. The pairwise and multiclass analysis is done in file-based and frame-based using Support Vector Machine (SVM) and k -Nearest Neighbours (k -NN) classifiers. Feature selection using Genetic Algorithm (GA) and ReliefF algorithm implementation to multiclass analysis are also investigated.

1.5 Thesis Outline

Chapter 1 introduced and presented the problems that arise in voice pathology detection including the research objective and scope.

Chapter 2 presents the literature review of related works and challenges in this study including database used, feature extraction and classification methods.

Chapter 3 explains the proposed methodology including the datasets used for the investigation, feature extraction, selection and classification methods implemented.

Chapter 4 discusses the results and evaluates the performance of the developed system in pairwise and multiclass for file-based and frame-based analysis.

Chapter 5 concludes the findings and research possibility to its application in the system and future development.

CHAPTER 2 : LITERATURE REVIEW

2.1 Introduction

Automatic systems have been designed based on signal processing or machine learning and data mining techniques for the detection of disease-related to vocal fold abnormalities (Amato et al., 2008). Figure 2.1 shows a general block diagram for vocal fold pathology detection, which includes voice signal pre-processing, feature extraction and classification. Feature extraction involved converting speech samples sequence into a set of observation vectors which represent events in a probabilistic space over which classification is performed (Saldanha, Ananthakrishna, & Pinto, 2013).



Figure 2.1 The general block diagram for vocal fold pathology detection.

This chapter explores the vocal fold pathology, databases, previous related works using wavelets and other feature extractions, feature optimization and classification methods for vocal fold pathology detection. The challenges, advantages and limitation of the application also been discussed.

2.2 Vocal Fold Pathology

Pathological disorders in the larynx can affect the speech signal quality and cause changes in the acoustic voice signal in the form of articulation (dysarthria), hoarseness (disruption of phonation), resonance, loss of power, changes in the pitch,

additional noises and, etc. (Nayak & Bhat, 2003; Vikram & Umarani, 2013; American Speech Language Hearing Association, 2020). The presence of surface irritation, tissue infection, mechanical stress, systemic alterations, tissue changes, muscular and neurological abnormalities can induce the vocal fold pathologies or voice disorders. These disorders known to influence the vocal folds during the phonation and may have a varied effect on vocal folds, as their capacity to close properly deteriorate (Ankışhan & İnam, 2021). Therefore, vibration and acoustic noise in vocal folds varies according to the type of pathologies. Some of the vocal fold pathologies are described in Table 2.1 (A. Rosen & Blake Simpson, 2008; Harar et al., 2018; Islam et al., 2020; American Speech Language Hearing Association, 2020; Johns Hopkins Medicine, 2021).

Table 2.1 Description of some vocal fold pathologies.

Vocal Fold Pathology	Description
Vocal nodules (callouses)	<ul style="list-style-type: none"> - Superficial scratches on the lining of the vocal folds usually occur in both vocal folds (middle) - Characterized as bilateral vocal fold lesions (fairly symmetric) - Difficult to produce higher range vocal pitches - Jitter (Perturbations or variations in frequency) and shimmer (amplitude) may be increased - Voice are breathy, low and hoarse
Vocal polyp (blister)	<ul style="list-style-type: none"> - Usually occurs on one vocal fold. - Firm or soft similar to a blister, and it is sometimes red due to haemorrhage. - Voice quality, loudness, pitch and ability to sustain voicing are evaluated. - Voice are breathy, low and hoarse, “rough” or “scratchy” voice
Vocal cyst	<ul style="list-style-type: none"> - A sac-like structure deeper in the vocal fold typically white or yellow in color under the vocal folds (superficial lining) - the mucosal wave reduced at the lesion - Problems with the pitch of the voice (tired voice), worsen after speaking with increased volume or for long periods of time
Vocal fold paralysis	<ul style="list-style-type: none"> - Vocal cords do not close or open properly (One or both folds) - Voice can be weak, swallowing trouble, liquids/food can slip into the trachea and lungs, and may choke or cough when they eat. - Noisy breathing and breathy voice, hoarseness, vocal pitch loss and inability to speak loudly (need to take frequent breaths and frequent throat clearing while speaking)
Laryngitis	<ul style="list-style-type: none"> - Inflammation of the vocal cords that can causes a raspy or hoarse voice

2.3 Databases

Voice pathology database becomes a crucial role in the field of pathological speech analysis since those voices become a viable procedure in clinical assessment for early detection of vocal fold pathology. Table 2.2 shows some of the databases developed to assist the researchers in the testing of non-invasive approaches for vocal fold pathology detection.

Table 2.2 Voice pathology database.

Database	Original Samples		Recording contents / Spoken words
	Healthy	Pathological	
Massachusetts Eye and Ear Infirmary (MEEI), English	53	657	3s sustained Vowel /a/
Saarbrücken Voice database (SVD), Germany	687	1320	1-3s Vowels /a/, /i/, /u/ (spoken at low, high, normal, and low-high-low intonations)
Principe de Asturias (PdA), Spain	239	200	A sustained Spanish vowel /a/.
Czech Parkinsonian Speech (PARCZ)	52	57 (Parkinson's disease suffer from hypokinetic dysarthria)	- Vowel /a/ - 91 speech tasks by elder people (reading text diadochokinetic tasks, free speech, maintained vowels, etc.)
Arabic Voice Pathology Database (AVPD)	188	178 (Sulcus, Nodules, Cyst, Paralysis and Polyp)	- sustained vowels, paragraphs and words - native speakers (Arabic language)
Advanced Voice Function Assessment Databases (AVFAD), Portugal	363	346	- Sustaining vowels /a/, /i/, /u/ - spontaneous speech, reading a phonetically balanced text and six CAPE-V sentences
RusDS, Belarusian Republican Center of Speech, Voice and Hearing Pathologies Russia	500	500 (Vocal Fold Paralysis)	1 s vowel /a/
Multiquality Database, General Hospital "Doctor Negrín" in Gran Canaria Spain	85	57 (Chronic Laryngitis, Hypofunction, Hyperfunction, Sessile Polyp, Vocal Fold Paralysis, Carcinoma, Reinke's Edema, Pedunculated Polyp, Adult Papiloma, Ulcer, Vocal Folds Nodule)	The five Spanish vowels (/a/, /e/, /i/, /o/, /u/ in the International Phonetic Alphabet - each vowel pronounced for approximately 2 seconds in a sustained way separated by silences