AN ENHANCED WAVELET NEURAL NETWORK MODEL FOR EPILEPTIC SEIZURE DETECTION AND PREDICTION

by

LAI KEE HUONG

Thesis submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

March 2016

ACKNOWLEDGMENTS

First and above all, I praise The Almighty God for the grace and mercy He bestowed on me. I thank Him for His wise guidance and rich blessing.

The journey of my postgraduate study and the completion of this thesis would not have been possible without the invaluable assistance and support from several institutions and individuals. I would like to take this opportunity to express my profound gratitude and indebtedness to them. The research work has been made more rewarding owing to the following people and institutions:

Prof. Zarita Zainuddin, my main supervisor, for her excellent supervision during my study. She has been a true mentor in every sense of the word and this thesis would not have been completed in the absence of her continued patience, inspiration, and guidance. I am eternally grateful for the opportunity provided by her to explore the fascinating realm of artificial intelligence.

Dr. Ong Pauline, my field supervisor, for her gracious generosity in sharing her vital research skills and methodologies. Her innovative ideas and suggestions have also made the research work more fruitful.

The Ministry of Higher Education, Malaysia, for the generous financial support in the form of Fundamental Research Grant Scheme (FRGS) that funded a significant portion of the research. Part of the research works was supported by this funding to cover conference presentations and publication fees.

The Institute of Postgraduate Studies (IPS), Universiti Sains Malaysia (USM), for awarding the prestigious USM Fellowship Scheme during the first three years of my study. The financial assistance provided has been enormously helpful during the course of my study.

The School of Mathematical Sciences, Universiti Sains Malaysia, for entrusting me with the position of tutor, as part of the school's duty under the requirement of the USM Fellowship Scheme. The experience gained in the teaching and learning process over the course of six semesters has played a key role in showing the way into a university career.

My primary and secondary schools' science and mathematics teachers, who have been a great source of inspiration over the years. They taught me the beauty and elegance of science and mathematics. I owe my love for mathematics to this group of diligent and dedicated educators. They are Mdm. Anna Ting, Mdm. Lai Jia Ling, Mdm. Lau Ming Hung, Mdm. Lee Yun Chin, and Mdm. Nora Ngo.

My friends, who have stood by me through thick and thin. I treasure the friendship and value these few special individuals for their long-term moral support and understanding. To Anna Ling Nget Min, Huong Kai Hee, Timothy Tie Dong Bing, Ting Lee Kang, Ting Siew Hiong, Ung Yew Jye, Wong Hua Jing, Wong Yu Chuan, and all my other friends, I thank you for all the laughter and joy in my life.

Last, but by no means least, I am deeply indebted to my parents and my siblings, for their unconditionally love and everlasting support. Mathematics has been my favorite subject for as long as I can remember, and I am thankful that I can further my study in a field that I am passionate about.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

LIST OF ABBREVIATIONS

ABC	Artificial Bee Colony
ABHS	Adjustable Bandwidth Harmony Search
ACO	Ant Colony Optimization
ADALINE	Adaptive Linear Neurons
AI	Artificial Intelligence
AIRS	Artificial Immune Recognition Systems
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANNs	Artificial Neural Networks
ApEN	Approximate Entropy
AR	Autoregressive
ARMA	Autoregressive Moving Average
BBA	Binary Bat Algorithm
BA	Bat Algorithm
BP	Backpropagation
BW	Bandwidth
CFO	Central Force Optimization
CHS	Chaotic Harmony Search
CNN	Combined Neural Network
CWT	Continuous Wavelet Transform
DB	Daubechies

DBHS	Discrete Binary Harmony Search
DFNN	Dynamic Fuzzy Neural Network
DWT	Discrete Wavelet Transform
ECG	Electrocardiography
EEG	Electroencephalography
EMG	Electromyography
ELM	Extreme Learning Machine
FA	Firefly Algorithm
FBI	Federal Bureau of Investigation
FCM	Fuzzy C-Means
FCMHS	Fuzzy C-Means Harmony Search
FFNNs	Feedforward Neural Networks
FPR	False Positive Rate
FT	Fourier Transform
GA	Genetic Algorithm
GHS	Global Harmony Search
GRASP	Greedy Randomized Adaptive Search Procedure
GAS	Gravitational Search Algorithm
HKA	Harmony K-Means
HM	Harmony Memory
HMCR	Harmony Memory Consideration Rate
HMS	Harmony Memory Size

HS	Harmony Search
IABHS	Improved Adaptive Binary Harmony Search
IBPSO	Improved Binary Particle Swarm Optimization
ICA	Independent Component Analysis
IHS	Improved Harmony Search
KM	K-Means
LDA	Linear Discriminant Analysis
KMHS	K-Means Harmony Search
MA	Memetic Algorithm
MADALINE	Multiple ADALINEs
MCP	McCulloch-Pitts
MEs	Mixture of Experts
MLPs	Multilayer Perceptrons
MRA	Multiresolution Analysis
MSFCM	Modified Point Symmetry-Based Fuzzy C-Means
NI	Number of Improvisation
OPF	Optimum-Path Forest
PAR	Pitch Adjusting Rate
PCA	Principal Component Analysis
PNNs	Probabilistic Neural Networks
PSD	Power Spectral Density
PSO	Particle Swarm Optimization

QMF	Quadrature Mirror Filters
RACO	EnRiched Ant Colony Optimization
RBFNs	Radial Basis Function Networks
ROC	Receiver Operating Characteristic
RLS	Recursive Least-Squares
RNNs	Recurrent Neural Networks
SA	Simulated Annealing
SBFS	Sequential Backward Floating Search
SCWN	Self-Constructing Wavelet Network
SFFS	Sequential Forward Floating Search
SGHS	Self Adaptive Global Harmony Search
SHS	Self Adaptive Harmony Search
SNNs	Spiking Neural Networks
SP	Saturation Parameter
STFT	Short Time Fourier Transform
SVMs	Support Vector Machines
T2FCM	Type-2 Fuzzy C-Means
T2FCMHS	Type-2 Fuzzy C-Means Harmony Search
TS	Tabu Search
TSP	Traveling Salesman Problem
WFT	Windowed Fourier Transform
WHO	World Health Organization

WT	Windowed Transform
WNNs	Wavelet Neural Networks

SUATU MODEL RANGKAIAN NEURAL WAVELET YANG DITAMBAH BAIK UNTUK PENGESANAN DAN RAMALAN SERANGAN EPILEPSI

ABSTRAK

Epilepsi merupakan suatu penyakit neurologi yang sangat lazim dan ditakuti orang ramai. Banyak kajian telah dibuat untuk membangunkan pengelas automatik yang dapat memberikan ketepatan yang lebih tinggi. Pengelas automatik ini dapat membantu doktor dalam mengenali pelbagai segmen isyarat electroencephalography (EEG) yang berbeza. Dalam kerja penyelidikan ini, suatu model rangkaian neural wavelet (RNW) telah dicadangkan bagi tujuan pengesanan dan ramalan serangan epilepsi. Arkitektur dan konfigurasi RNW dapat ditambah baik menggunakan pendekatan metaheuristik. Khususnya, algoritma carian harmoni (CH) digunakan dan diterapkan dalam proses pembelajaran RNW. Tesis ini mengandungi tiga sumbangan utama. Pertama, algoritma CH digunakan dalam proses pemilihan fitur. Algoritma CH, yang pada asalnya digunakan untuk menyelesaikan masalah pengoptimuman yang melibatkan nombor nyata, telah diubah suai dan digunakan dalam proses pemilihan fitur yang melibatkan nilai binari. Di samping meringkaskan arkitektur rangkaian, penurunan dalam bilangan fitur turut dapat mengurangkan kos komputasi. Kedua, algoritma CH digunakan untuk menentukan lokasi vektor anjakan dalam neuron tersembunyi RNW. Suatu set vektor anjakan yang baik secara tidak langsung berupaya meningkatkan kecekapan proses pembelajaran RNW. Untuk mencapai matlamat ini, algoritma CH dihibridkan dengan algoritma c-min kabur jenis kedua. Ketiga, algoritma CH diterapkan dalam algoritma pembelajaran RNW. Algoritma CH secara khususnya digunakan untuk menentukan nilai pemberat sinaptik dan bias. Strategi inisialisasi memori harmoni dan improvisasi yang baru diperkenalkan dalam algoritma pembelajaran CH yang dicadangkan. Keberkesanan ketiga-tiga kaedah penambahbaikan tersebut diuji dengan sepuluh set data pembelajaran mesin UCI. Simulasi awal melaporkan bahawa kaedah hibrid menunjukkan prestasi yang lebih unggul berbanding dengan algoritma konvensional yang standard. Selain itu, model RNW yang dilatih dengan CH melaporkan keputusan yang setanding dengan model RNW yang dilatih menggunakan algoritma metaheuristik yang lain. Model RNW yang ditambah baik dalam tiga aspek yang berbeza kemudiannya diuji dalam dua aplikasi dunia sebenar, iaitu pengesanan dan ramalan serangan epilepsi. Kaedah transformasi wavelet diskrit digunakan untuk memproses isyarat EEG bagi menghasilan kumpulan pekali wavelet yang berbeza, berpadanan dengan jalur frekuensi masing-masing. Keputusan simulasi melaporkan bahawa model RNW yang ditambah baik menunjukkan prestasi yang lebih baik jika dibandingkan dengan kaedah pembelajaran mesin lain yang dilaporkan dalam literatur. Keputusan ini menunjukkan potensi penggunaan dan pelaksanaan model RNW yang dicadangkan dalam bidang epileptologi.

AN ENHANCED WAVELET NEURAL NETWORK MODEL FOR EPILEPTIC SEIZURE DETECTION AND PREDICTION

ABSTRACT

Epilepsy is a very common and much-feared neurological disorder. Much research has been done in developing better automated classifiers with higher accuracy that can help clinicians identify the different segments of electroencephalography (EEG) signals. In this research work, an enhanced wavelet neural network (WNN) model is proposed for the purpose of epileptic seizure detection and prediction. The architecture and configuration of WNNs can be further enhanced using metaheuristic strategies. Specifically, the harmony search (HS) algorithm is employed and incorporated in the learning of WNNs. The contribution of this thesis is threefold. Firstly, the HS algorithm is used in the feature selection stage. The HS algorithm, which is originally used for optimization problems involving real numbers, is modified and employed in the task of feature selection, which involves binary values. Apart from simplifying the network architecture, the reduction in the number of features also reduces computational cost. Secondly, the HS algorithm is employed to find the translation vectors of the hidden nodes of WNNs. A good set of translation vectors will indirectly increase the efficiency of the learning process of WNNs. To achieve this goal, the HS algorithm is hybridized with the type-2 fuzzy c-means clustering algorithm. Thirdly, the HS algorithm is incorporated in the learning algorithm of WNNs. In particular, the HS algorithm is used to determine the synaptic weights and bias terms of WNNs. Novel harmony memory initialization and improvisation strategies are incorporated in the proposed HS-based learning algorithm. The effectiveness of the three aforementioned improved methods are first tested using ten sets of UCI machine learning data sets. The preliminary simulations report that the hybridized methods give superior performance than the conventional stand-alone algorithms. Also, WNNs models that are trained using the HS algorithm and other metaheuristic approaches report comparable results. The WNNs models with enhancements in three different aspects are then tested using two real world applications, namely in the tasks of epileptic seizure detection and prediction. The discrete wavelet transform (DWT) method is used to pre-process the EEG signals to yield different groups of wavelet coefficients, which correspond to different frequency sub-bands. Simulation results show that the enhanced WNN model outperforms most of the other machine learning methods reported in the literature. This suggests the potential usage and implementation of the developed classifiers in the field of epileptology.

CHAPTER 1
INTRODUCTION

1.1 Preliminaries

Epilepsy is a common neurological disease, affecting approximately 50 millions people worldwide (WHO, 2015). This medical disorder is characterized by the occurrence of recurrent seizures. The development of better expert systems in the diagnosis of epileptic seizure is hence, of utmost importance. Automated classifiers with higher accuracy can help clinicians identify and evaluate the different segments of electroencephalography (EEG) signals. Artificial neural networks (ANNs), with enhancement using metaheuristic methods in various learning aspects, are popular mathematical models that are used for this purpose.

In the field of biomedical engineering, the classification of biomedical signal is an important decision-making task. To further clarify the title of the thesis, the three terminologies, namely classification, detection, and prediction are first made clear. The term *classification* refers to the task of classifying a given EEG signal into one of the many subclasses. Both epileptic seizure detection and prediction fall into the scope of classification. The task of epileptic seizure detection aims at distinguishing the normal (interictal) EEG signals and the epileptic (ictal) EEG signals. The term *detection* that is used in this context refers to the job of detecting or identifying the abnormal or epileptic EEG signals from the normal EEG signals. On the other hand, the task of epileptic seizure prediction is basically a classification task as well, only this time, it aims at differentiating between interictal and pre-ictal (before seizure) EEG signals. The term *prediction* is used because the ANNs models are used to predict the occurrence of impending seizure attacks. If a given EEG signal is classified as a pre-ictal signal, then it implies that the mathematical model predicts that a seizure attack will occur within a given time frame.

The very first chapter of this thesis gives a brief introduction of the medical condition of epilepsy. After reviewing the works that have been done in the domain of epileptic seizure, the motivations are given regarding the use of automated classifiers in the tasks of epileptic seizure detection and prediction. Next, the problem statements, objectives, and significance of the research are given. Lastly, an overview of the outline and organization of the thesis is presented.

1.2 Epilepsy

After stroke, epilepsy is the second most common neurological disorder. The disease manifests itself in the form of epileptic seizure, caused by the excessive firing of neurons in the brain (WHO, 2015). Some cases of epilepsy can be treated or even cured by means of medication and surgery. However, some epilepsy cases are deemed intractable. Patients diagnosed with epilepsy suffer from both economic disadvantages and social discrimination.

Much effort and research have been done in the field of epileptic seizure detection (Orosco et al., 2013) and prediction (Carney et al., 2011; Ramgopal et al., 2014). The interdisciplinary works in epileptic seizure detection and prediction involve collaborative efforts from epileptologists, biomedical engineers, computer scientists, and mathematicians. Different types of expert systems have been developed using various promising feature extraction techniques and powerful artificial intelligence-based models (Acharya et al., 2013). Among the models that are reported in the literature include artificial neural networks (ANNs), mixture of experts (MEs), and support vector machines (SVMs). Some metaheuristic and intelligent methods, such as genetic algorithm (GA) and particle swarm algorithm (PSO) are embedded and integrated in the learning of these models. Mathematical models with higher accuracy would benefit the epileptic patients and medical community at large in accident prevention, as well as the development of closed-loop seizure prediction warning systems.

1.3 Motivation

Due to the time constraint and inconsistency of human judgment, the use of automated classifiers and expert systems is invaluable in the decision making process in the medical field. Apart from reducing medical expenditure, the use of such artificial intelligence-based approach can also save time. In addition, the feasibility and practicality of using neural network models in facilitating classification and pattern recognition problems are corroborated by the fact that these mathematical models are not affected by human fatigue, emotional states, and other undesirable factors (Micheli-Tzanakou, 1995).

Neural networks models that can yield classification with high accuracy are desirable because they are utilized widely in the decision-making process. In medical domain, mathematical models are used to classify biomedical signals. Furthermore, these expert systems are used to determine whether a given tissue culture contains cancerous cells. As such, high classification accuracy is a very important criterion of efficient neural networks models. ANNs models with very low classification error and false positive rate (FPR) are imperative in medical diagnosis. In this research work, the wavelet neural networks (WNNs) models are considered. To improve the performance of the WNNs models, the enhancements are accomplished by integrating the metaheuristic harmony search (HS) algorithm in three different learning aspects, namely feature selection, cluster initialization, and supervised learning.

The main task of the WNNs models developed in this work is to differentiate between two classes of electroencephalography (EEG) biomedical signals. The objective of epileptic seizure detection is to differentiate between interictal and ictal EEG signals. Epileptic patients need to be monitored for pre-surgical evaluation purpose. Scrutinizing all the EEG data recorded for days or even weeks manually is a painstaking and meticulous task. As such, a more realistic solution is the development of automated classifiers that can perform such tasks at high speed with great accuracy.

On the other hand, the ultimate goal of epileptic seizure prediction is to distinguish between interictal and pre-ictal EEG signals. By identifying the pre-ictal portion of the biomedical signals, an alarm can be issued to alert the patients of impending seizure attacks. The realization of such classifiers also paves the way for better and more efficient epileptic seizure closed-loop intervention strategies, via drug administration and seizure warning devices similar to vagus nerve stimulator.

The development of such detection and prediction models will not only improve the quality of life of epileptic patients, but it also benefits the medical community at large.

1.4 Problem Statements

The emphasis of this work is the development of an enhanced WNN model. The improvements are accomplished in three aspects, namely feature selection, initialization of the translation vectors, as well as learning algorithm. All these enhancements are aimed at designing artificial intelligence-based classifiers that are able to make fast and accurate judgments with high classification accuracy.

The three main problem statements addressed by this research are:

(i) Finding the optimal feature subset of reduced size

In many scientific and engineering applications, a huge amount of data is generated and collected. Before the data are fed into the input layer of ANNs, they need to undergo some pre-processing stages, such as feature extraction and selection. This is done to eliminate noise and outliers, which are common due to human error, erroneous measurement, and calibration error. By eliminating irrelevant and redundant features, a good feature selection algorithm is also able to reduce the dimension of input data, thereby simplifying network topology and training time. An efficient feature selection algorithm (Chandrashekar and

Sahin, 2014) is important so that the number of input nodes of the ANNs models used is as minimum as possible. The reduction of input nodes implies a simpler network architecture.

(ii) Determining the locations of the best translation vectors

Central to the topic of discussion in Uykan et al. (2000) and Guillén et al. (2005) is the issue of selecting the most optimal set of centers or translation vectors for radial basis function networks (RBFNs). An efficient clustering algorithm that can find the best locations of the centers is crucial because the performance of RBFNs is highly dependent on the selection of these centers. A bad choice of centers will undoubtedly affect the subsequent learning process where the values of synaptic weights are determined. Similar to the network design of RBFNs, WNNs also employ localized activation functions in the hidden nodes, where the locations of translation vectors need to be determined beforehand. On the contrary, a good set of translation vectors can increase the generalization capability of WNNs (Ong and Zarita, 2016).

(iii) Devising an effective learning algorithm that can find the optimal weight parameters

Carefully examining the network architecture and learning algorithm of the conventional multilayer perceptrons (MLPs), which are one of the earliest ANNs models, reveal some drawbacks that limit their use and application. The limitations of the MLPs models include the use of global activation functions, failure to converge in the case of highly nonlinear data, tendency of getting trapped at local minima, and time-consuming training process (Ham and Kostanic, 2000; Oysal et al., 2005). The use of global functions, such as the sigmoid functions in the hidden nodes of MLPs is undesirable because the functions span over a wide range and they will activate all the input fed to the ANNs. Hence, the use of WNNs with localized wavelet functions, coupled with metaheuristic method in the learning process is proposed in this work. The metaheuristic algorithm is used to determine the optimal values of the weight parameters and bias terms.

In light of this, an enhanced WNN model is proposed to address the aforementioned shortcomings. Unlike global activation functions, the localized wavelet functions embedded in the hidden nodes of WNNs have finite support, which means that they will only activate a subset of input whose values are close to the translation vectors. Furthermore, to address the dreaded problem of solutions getting trapped at local minima, the evolutionary harmony search (HS) algorithm is incorporated in several aspects of the training process of WNNs. The HS algorithm is capable of finding near-optimal solutions by exploring the entire solution search space effectively within a reasonable amount of iteration and time.

1.5 Research Objectives

The main objective of this research is to develop an enhanced wavelet neural network (WNN) model, through the incorporation or hybridization of the metaheuristic harmony search (HS) algorithm, in three different aspects, namely feature selection, translation vectors initialization, and learning algorithm.

The objectives of this thesis are listed as follows:

(i) To devise an efficient feature selection algorithm by finding the optimal feature subset of smaller size so as to reduce the dimensionality of the input nodes and subsequently, the network topology and complexity. The decision of including or excluding a particular feature in the proposed feature subset is guided by a set of rules (governed by global and local search) that are embedded in the metaheuristic algorithm.

(ii) To develop a novel clustering algorithm that is able to locate the translation vectors of the hidden nodes of WNNs for the purpose of increasing the performance of WNNs in terms of accuracy and generalization capability. The factors that are taken into consideration when selecting the best translation vectors include the possible presence of outliers in the input data. Furthermore, a more flexible rule is introduced by using a fuzzy approach where a particular input datum can be assigned to all the cluster centers.

(iii) To improve the learning aspects of WNNs through the implementation of a new learning algorithm by integrating the metaheuristic approach to search for near-optimal solutions. The best values of the synaptic weight values and bias terms are determined using the iterative metaheuristic search. The metaheuristic approach used ensures that the entire solution space can be explored thoroughly during the initial exploration stage. This is followed by the local exploitation stage, where the solutions are fine-tuned gradually.

(iv) To demonstrate the effectiveness of the proposed enhanced WNNs in the binary classification tasks of epileptic seizure detection and prediction by designing powerful expert systems that can be used in clinical settings. The effectiveness and robustness of the proposed models are characterized by high classification accuracy, high sensitivity value, and low false positive rate.

1.6 Thesis Organization

This thesis is divided into five main sections – preliminaries, theoretical frameworks, contributions, applications, and concluding remarks. Each section is addressed in one or more chapters.

The first section presents the preliminary concepts of the research work. Chapter 1 gives a brief introduction of the medical condition of epilepsy. The motivation is then provided to highlight the importance of the research done in the domains of epileptic seizure detection and prediction. Additionally, problem statements, main objectives, and the significance of the research are given.

The second section concerns the theoretical frameworks that are used in this work. Chapter 2 reviews the concepts of wavelet theory, discrete wavelet transform (DWT), artificial intelligence (AI), artificial neural networks (ANNs), wavelet neural networks (WNNs), and the metaheuristic harmony search (HS) algorithm. In particular, the network architecture, parameters initialization, and learning algorithm of WNNs are discussed. This is followed by the discussion of the history, motivation, and development of the HS algorithm. Next, the algorithm of the standard HS algorithm is presented, followed by the literature review of the variants of the HS algorithms, developed for optimization problems that deal with real, discrete, and binary decision variables. Additionally, the applications of the HS algorithm are illustrated.

The third section of the thesis includes three chapters that focus on the three main contributions of this research. Chapter 3 begins with the review of the existing feature selection methods, which are filter, wrapper, and hybrid approaches. Then, an enhanced feature selection algorithm is proposed. The effectiveness of the proposed algorithm is verified using the UCI benchmark data sets. Chapter 4 highlights the use of clustering algorithms in the initialization of translation vectors of the hidden nodes of WNNs. The standard k-means and fuzzy c-means clustering algorithms are first presented before the proposal of the novel hybridized clustering algorithm. Numerical simulations are performed on the same benchmark data sets to validate the robustness of the proposed method. Chapter 5 examines the existing learning algorithms used for ANNs and in particular, WNNs. The idea of the incorporation of the metaheuristic HS algorithm in the learning algorithm of WNNs is then presented. The method is tested using ten UCI data sets.

The fourth section investigates the feasible application of the enhanced WNN model in the classification tasks of epileptic seizure detection and prediction. Chapter 6 consists of two main parts. The first part begins with the explanation of some medical terminologies, such as epilepsy, epileptic seizure, and electroencephalography (EEG) signals. The literature review on the state of the art of the methods used and reported in the task of epileptic seizure detection is surveyed. Next, the methodology and experimental design are given. These include the stages of data acquisition, feature extraction, feature selection, and the classification using the enhanced WNNs models. The results obtained are evaluated and the relevant discussion is provided. The second part of chapter 6 presents a more challenging classification task of epileptic seizure prediction. The main difference between epileptic seizure detection and prediction are first made clear. Some technical terms used in the research of seizure prediction, such as pre-ictal period and seizure occurrence period are defined. The existing methods used for the purpose of epileptic seizure prediction are covered in the literature review. Next, the experimental design, results, and discussion are reported.

The last section, namely section five, is a write up on some concluding remarks. Chapter 7 concludes the main findings of this work and gives some suggestions for future work.

CHAPTER 2

WAVELET NEURAL NETWORKS AND HARMONY SEARCH ALGORITHM

2.1 Introduction

This chapter presents the theoretical frameworks that are used in the research work. The first part of this chapter gives an exposition of the many facets of the fascinating realm of wavelet theory, as well as several sub-disciplines that stem from this field of study. A brief introduction on wavelets is first reviewed, followed by the exploration of their properties. Next, a timeline on the development of the field of wavelet theory is presented. To appreciate the advantages offered by wavelet analysis, the limitations of Fourier transform (FT) are examined. Then, the distinction between discrete wavelet transform (DWT) and continuous wavelet transform (CWT) are explained. The broad spectrum of the applications of wavelets is also given. The notion of artificial intelligence (AI) is then presented. This is followed by the discussion on the history, properties, and applications of artificial neural networks (ANNs). The next part of this chapter introduces the wavelet neural networks (WNNs) models, where the emphasis of discussion is given to the network topology, parameterization, and applications.

The second part of this chapter presents the theoretical framework of the harmony search (HS) algorithm. The concept of metaheuristic algorithms is first explained, followed by the discussion of the standard HS algorithm. The idea and motivation of the HS algorithm are presented. To fully comprehend how the algorithm works, its working mechanism is detailed using a simple illustrative example. In addition, the convergence of the HS algorithm is presented to demonstrate that sub-optimal solutions are always guaranteed. Some of the notable variants of the HS algorithms reported in the literature are covered and discussed in the subsequent subsections. The applications of the HS algorithm are also given. The use of various metaheuristic methods in the domain of machine learning is reviewed at the end of the chapter.

2.2 Wavelet Theory

Wavelets are versatile mathematical tools that are at the heart of many applications such as electrical engineering and quantum physics. The fascinating properties of wavelets, such as finite support, perfect symmetry, and high smoothness, make wavelet analysis an ideal choice in the disciplines of signal processing, image compression, and noise removal. Discrete wavelets are used extensively in discrete wavelet transform, whereas continuous wavelets are employed as activation functions in the hidden nodes of wavelet neural networks (WNNs).

2.2.1 Introduction to Wavelets

Etymologically, the mathematical lexicon "wavelet" is derived from the French word *ondelette*, which means "small wave". The terminology was first coined by French geophysicist Jean Morlet and Croatian physicist Alexander Grossmann in their seminal works in wavelet analysis (Grossmann and Morlet, 1984).

Mathematically speaking, a wavelet is a special type of function that meets certain criteria. As shown in Figure 2.1, unlike sinusoidal functions (e.g., the wave-like sine and cosine trigonometric functions), wavelets are localized functions where they decay rapidly toward zero as their limits approach infinity. In other words, wavelets have finite energy, whereas sinusoids have infinite energy.

A mother wavelet must meet the following three conditions:

(i) The integral of the wavelet has zero mean.

$$\int_{-\infty}^{\infty} \psi(t)dt = 0. \tag{2.1}$$

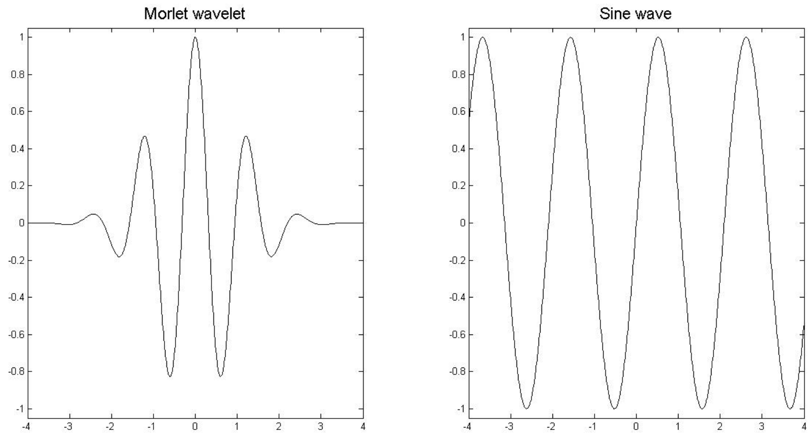

Figure 2.1: Comparison of Morlet wavelet function and sine wave function

(ii) The integral of the square of the wavelet is unity.

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1. \tag{2.2}$$

(iii) Admissibility condition.

$$C_\psi = \int_0^{\infty} \frac{|\Psi(\omega)|^2}{\omega} d\omega < \infty, \tag{2.3}$$

where $\Psi(\omega)$ is the Fourier transform of $\psi(t)$, given by the following formula:

$$\Psi(\omega) = \int_{-\infty}^{\infty} \psi(t) e^{-it\omega} dt. \tag{2.4}$$

The admissibility condition implies that at zero frequency, the Fourier transform of $\psi(t)$ vanishes, as shown in the following equation:

$$|\Psi(\omega)|^2_{\omega=0} = 0. \tag{2.5}$$

From a mother wavelet $\psi(t)$, a family or a series of wavelet functions, termed daughter wavelets, can be generated via translation and scaling parameters. The translation parameter shifts the center of the wavelet, whereas the scaling parameter changes the appearance of the wavelet, either by stretching or shrinking the function. The daughter wavelets are generated using the following formula:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right), a, b \in \mathbb{R}, a \neq 0, \tag{2.6}$$

where a and b are the scaling and translation parameters, respectively. The normalization term $\frac{1}{\sqrt{|a|}}$ is introduced to ensure that the energy $\|\psi_{a,b}(t)\|$, is independent of the values of a and b.

2.2.2 Properties of Wavelets

Wavelets are regarded as one of the most promising tools in many useful applications. Some of their most significant and notable properties are listed as follows:

(i) Vanishing moments

If a wavelet scaling function is able to generate polynomials up to degree $p-1$, then the wavelet function is said to have p vanishing moments. A wavelet that has a higher number of vanishing moments can represent more complicated functions.

(ii) Compact support

A wavelet function is non-zero only on a finite or limited range of its domain. Outside this range or interval, the wavelet function decays to zero. This property makes continuous wavelet functions a perfect choice as the transfer functions of hidden nodes, where only a limited amount of input values will be activated.

(iii) Regularity

The regularity of a wavelet is closely related to its vanishing moments. The fast decaying characteristic of wavelet is attributed to this property.

(iv) Smoothness

The smoothness of a wavelet is determined by its vanishing moments. A smoother wavelet with more negligible wavelet coefficients whose values are close to zero is desirable, as this property is imperative in the task of image compression.

(v) Symmetry

A wavelet that is symmetrical in shape is essential because it is used as the building block of bases. In signal processing, the problem of phase distortion is caused by asymmetrical wavelet.

2.2.3 Timeline of Wavelets

The subject area of wavelet analysis has witnessed tremendous development and progress over the past few decades since its inception in the early 20th century. Table 2.1 highlights the major breakthrough in this field.

2.2.4 Fourier Transform

A complex mathematical problem at hand that is seemingly difficult to solve in its original setting could be solved seamlessly in another domain. This is the essence of the brilliant mathematical transformation. By transforming the question presented into a different domain, the solution can be obtained relatively easier, and in a faster way. The answer is then converted back to the original domain. Some common mathematical transformations are such as integration by parts and the Laplace transform. The former method transforms the antiderivative of a product of two functions into another antiderivative that can be solved easier, whereas the latter approach reduces a complicated differential equation into a simpler algebraic equation.

Another commonly used transformation is the Fourier transform (FT) that

Table 2.1: Timeline of the development and major breakthrough in wavelet analysis

1807	Jean-Baptiste Joseph Fourier presented the idea that any complex periodic function can be expressed as a sum of sine and cosine functions (also called prototype or basis functions).
1909	Alfréd Haar proposed the first wavelet. The Haar wavelet consists of a positive pulse and a negative pulse. The wavelet is discontinuous and non-differentiable in nature, and these shortcomings limit its application.
1930	Paul Lévy discovered that the Haar wavelets outperformed the Fourier basis functions. The observation was made during his Brownian motion research.
1946	Dennis Gabor noticed that Fourier transform could not provide sufficient information for time-frequency analysis; therefore, he suggested the use of a window function in short time Fourier transform (STFT).
1981	Jean Morlet and Alexander Grossmann suggested the use of different window functions to analyze signals at different frequency sub-bands. The technique was used to study seismic signals encountered widely in underground oil search operations.
1985	Yves Meyer reported the first smooth orthogonal wavelets that have better localization properties in both time and frequency domains.
1986	Stéphane Mallat developed the technique of multiresolution analysis (MRA), in which discrete signals are decomposed into several frequency sub-bands using lowpass and highpass filters. The theory of MRA is similar to quadrature mirror filters (QMF) in electronic engineering.
1987	Ingrid Daubechies laid the foundation for modern wavelet theory by introducing a family of Daubechies wavelets. These Daubechies wavelets are orthogonal and have compact support. They can be programmed and implemented easily using digital filters.

decomposes a given signal into its individual frequencies. The process is analogous to the process of breaking down a chord into its individual notes in the field of music theory. The FT is defined in terms of Fourier series. Given any 2π periodic function $f(x)$, it can be expressed as an infinite sum of sine and cosine trigonometric functions. Mathematically speaking, the Fourier series of a piecewise continuous function $f(x)$ on $[-\pi, \pi]$ is given by

$$a_0 + \sum_{n=1}^{\infty}(a_n \cos(nx) + b_n \sin(nx)), \tag{2.7}$$

where the coefficients a_0, a_n, and b_n are given by

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)dx, \tag{2.8}$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)dx, \ n = 1, 2, 3, ..., \tag{2.9}$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)dx, \ n = 1, 2, 3, ... \tag{2.10}$$

The relationship between a function $f(x)$ and its Fourier transform $\hat{f}(\omega)$ is given by

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\omega)e^{2\pi i\omega x}d\omega, \tag{2.11}$$

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i x\omega}dx. \tag{2.12}$$

The Fourier transform is used to pre-process signals to obtain information that is not readily available in the raw signals themselves. The raw signals are functions of time. The dependent variable, amplitude, is plotted against the independent variable, time. The resulting graph shows a time-amplitude representation of the signal under study. Nevertheless, the most useful and distinguished information of a signal is embedded in its frequency spectrum that tells which frequency sub-bands exist in the signal.

In the field of signal processing, all signals are divided into two categories, namely stationary and non-stationary. As the name implies, the statistical parameters of stationary signals remain constant over time. In contrast, non-stationary signals have different parameter values on each time interval. In spite of the remarkable success of FT in analyzing stationary signals, the method proves to be inadequate in studying non-stationary signals. The main limitation of FT is that it only provides the spectral or frequency contents of signals and no temporal information at all. In other words, one could not tell at what time the spectral components appear. This main drawback is explained using the following example taken from Polikar (1996).

Consider the stationary signal $x_1(t)$ and non-stationary signal $x_2(t)$ defined by the following functions:

$$x_1(t) = \cos(2\pi(10t)) + \cos(2\pi(25t)) + \cos(2\pi(50t)) + \cos(2\pi(100t)), \tag{2.13}$$

$$x_2(t) = \begin{cases} \cos(2\pi(10t)) & , \quad 0 \le t \le 0.3 \\ \cos(2\pi(25t)) & , 0.301 < t \le 0.6 \\ \cos(2\pi(50t)) & , 0.601 < t \le 0.8 \\ \cos(2\pi(100t)) & , \ 0.801 < t \le 1. \end{cases} \tag{2.14}$$

The plot of the stationary signal $x_1(t)$ and the non-stationary signal $x_2(t)$ are given in Figure 2.2 and Figure 2.3, respectively.

As shown in Figure 2.2, the four different frequency components that appear in the signal $x_1(t)$ at any given time are 10 Hz, 25 Hz, 50 Hz, and 100 Hz. On the other hand, the non-stationary signal $x_2(t)$ shown in Figure 2.3 is made up of four functions that appear in different time intervals. Each of the four segments contains only one frequency component. Observe that the values of the four frequency components are identical to those of the stationary signal.

The plots of the Fourier transform of the two signals are given in Figure 2.4

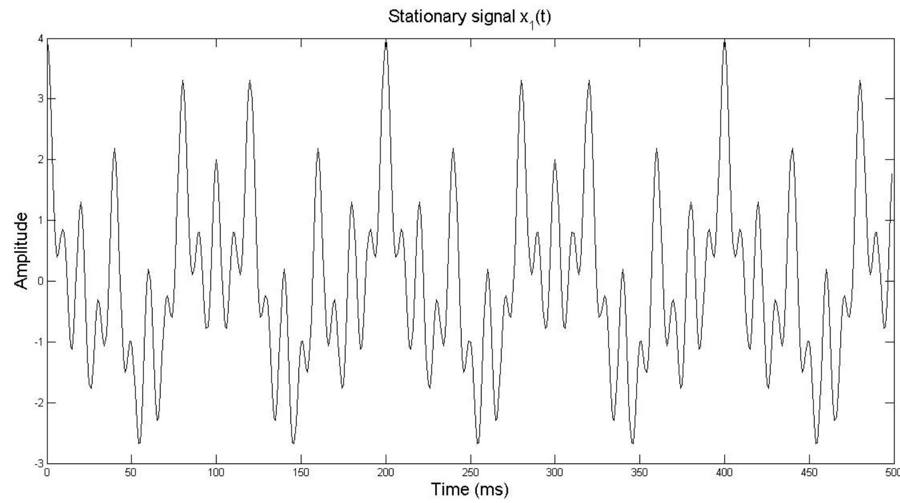

Figure 2.2: Stationary signal $x_1(t)$

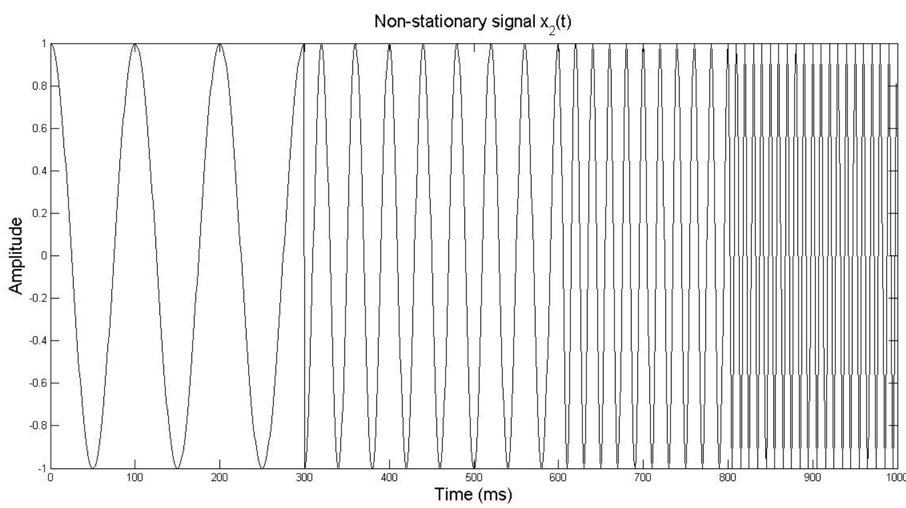

Figure 2.3: Non-stationary signal $x_2(t)$

and Figure 2.5, respectively. An interesting observation is made by comparing these two figures. Two totally different signals (one stationary and one non-stationary) yield similar Fourier transform plots. The peaks shown in the graphs correspond to the frequency components of the raw signals. This example demonstrates that the Fourier transform method can only provide spectral information (which frequencies exist in the signals), but not the temporal information (at what time the frequencies appear).

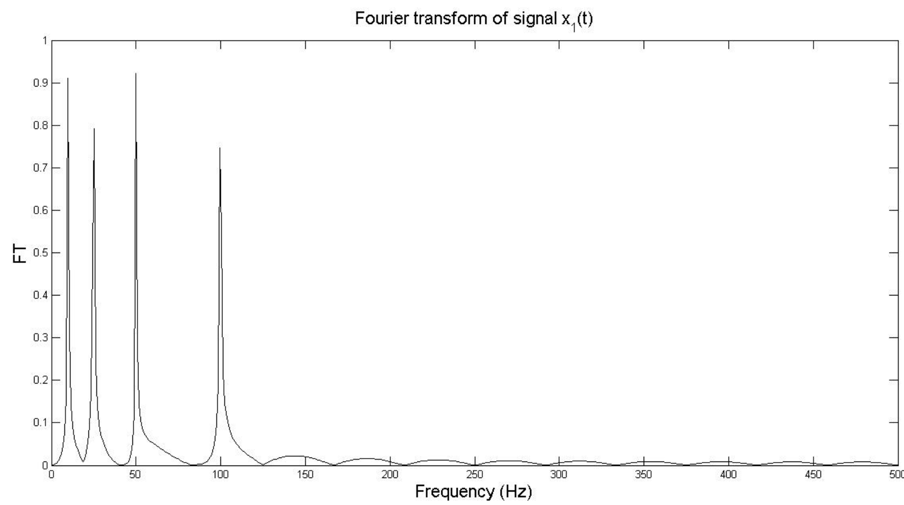

Figure 2.4: Fourier tranform of $x_1(t)$

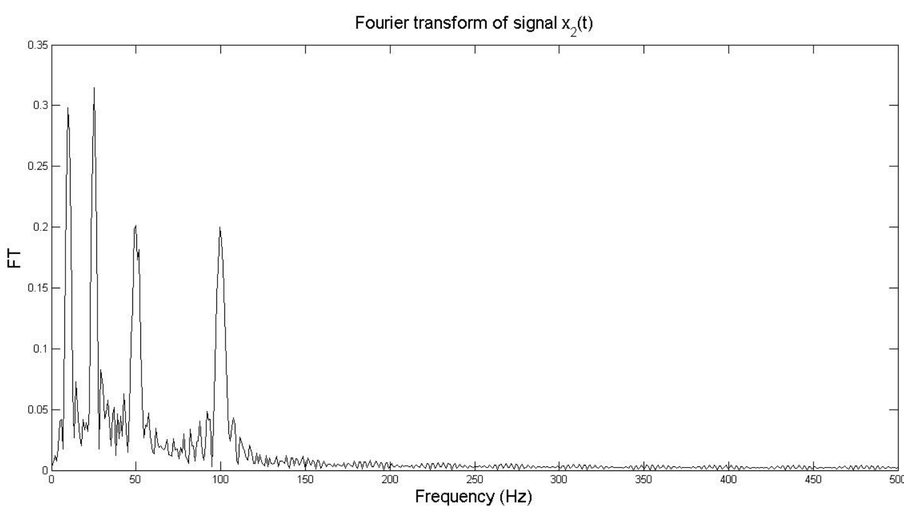

Figure 2.5: Fourier transform of $x_2(t)$

Realizing the pitfall of FT, another alternative, termed Windowed Fourier transform (WFT), or short time Fourier transform (STFT) is proposed by Gabor (1946). In this approach, a fixed-size window function is employed to analyze signals. The signal under study is divided into several smaller regions, where the signal in each smaller segment is assumed to be stationary. By shifting the window repeatedly over different regions of a signal, STFT is able to give a good time-frequency representation of the signal. STFT is essentially the same as FT, just that it differs in the use of a window function, as shown in the following equation

$$STFT\{x(t)\}(\tau,\omega) \equiv X(\tau,\omega) = \int_{-\infty}^{\infty} [x(t)w(t-\tau)]e^{-2\pi it\omega}dt, \tag{2.15}$$

where $x(t)$ is the signal, which is a function of time t, $w(t)$ is the window function, and τ is the time index.

Despite the ability of providing a time-frequency representation of the signal under study, a significant issue arises regarding its resolution. The width of the window function $w(t)$ plays a vital role in determining the time and frequency resolutions. To illustrate, a wide window, which covers a longer time interval, gives poor temporal resolution but good spectral resolution. On the other hand, a narrow window, which covers a shorter time interval, yields poor spectral resolution but good temporal resolution. This delicate issue concerning the trade-off between the temporal and spectral resolutions gives rise to the birth of wavelet transform (WT) and multiresolution analysis (MRA), where flexible window functions are used to study signals to preserve good temporal and spectral resolutions simultaneously. A complete and comprehensive analysis of non-stationary signals requires information of both spectral (frequency) and temporal (time) components.

2.2.5 Discrete Wavelet Transform

The main feature of wavelet transform (WT) is the ability to study a given signal using different scales, which is better known as multiresolution analysis (MRA). The comparison of FT, STFT, and WT is shown in the form of time-frequency plane in Figure 2.2.5. Although FT is able to capture good information in the frequency domain, the method gives poor localization in the time domain. On the other hand, the STFT uses a constant window function that results in a fixed time-frequency resolution. To overcome the limitations of the classical FT and STFT methods, the WT approach is the ultimate solution. WT offers an optimal compromise between the two spectral and temporal components. At high frequencies, WT employs a narrower window that gives good time localization but poor frequency localization. On the other hand, at low frequencies, the method uses a wider window that gives good frequency localization but poor time localization. The adaptive nature or flexibility of the time-frequency localization property makes WT an excellent tool in extracting useful information embedded in raw signals.

Wavelet transform can be accomplished in either continuous or discrete domain. The continuous wavelet transform (CWT) of a signal $f(t)$ is given by the following formula:

$$\begin{aligned} CWT_f(a,b) &= \int_{-\infty}^{\infty} \psi_{a,b}(t) * f(t)dt \\ &= \frac{1}{|a|^{-\frac{1}{2}}} \int_{-\infty}^{\infty} \psi\left(\frac{t-b}{a}\right) * f(t)dt, \end{aligned} \tag{2.16}$$

where a is the scaling parameters, b is the translation parameter, and ψ is the mother wavelet. The algorithm of CWT is given as follows:

(i) Choose a wavelet function $\psi(t)$ and compare it with a short segment taken from the start of the signal under study, $f(t)$.

(ii) Calculate the value of CWT. The value measures the similarity between the wavelet and the segment of the signal.

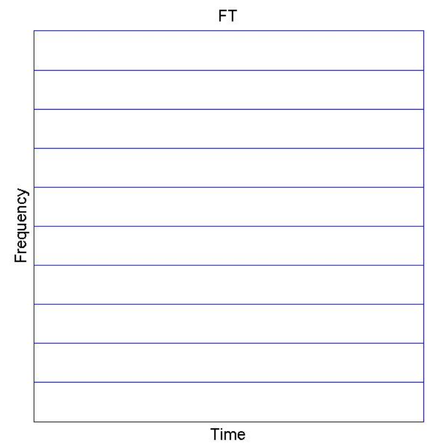

Figure 2.6: Time-frequency representation of Fourier transform

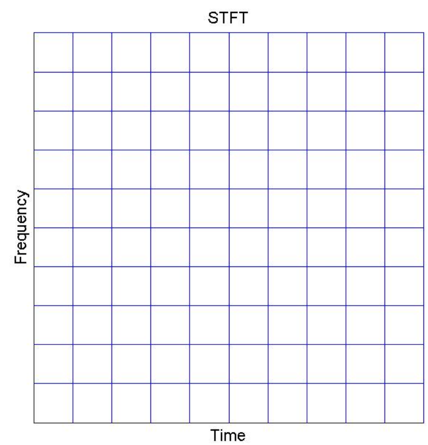

Figure 2.7: Time-frequency representation of short time Fourier transform

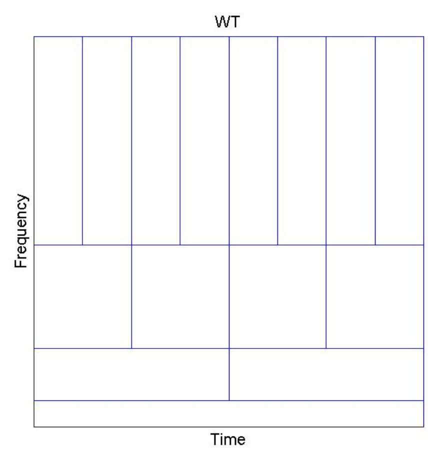

Figure 2.8: Time-frequency representation of wavelet transform

(iii) Shift the wavelet to the right using the translation parameter and repeat step (ii) until the entire signal is covered.

(iv) Choose a scaling parameter for the wavelet function. Repeat steps (i) to (iii).

(v) Repeat steps (i) to (iv) for all the different values of scaling parameter.

It is noticed that for CWT, the values of scaling and translation parameters are varied continuously over real numbers. As such, the CWT will generate a huge amount of wavelet coefficients, which is not only redundant, but the process is time-consuming and computationally costly. This drawback has led to the development of a more feasible discrete wavelet transform (DWT) approach, where only discretized values of the parameters are considered.

In DWT, the values of translation and scaling parameters are sampled discretely. The values of a and b are chosen in the following ways:

$$a = a_0^j, \forall j \in \mathbb{Z}, \tag{2.17}$$

$$b = k a_0^j b_0, \forall j, k \in \mathbb{Z}, \tag{2.18}$$

where $a_0 > 1$ and $b_0 \neq 0$ are the dilated and translated steps, respectively. The family of wavelets generated is given by

$$\psi_{j,k}(t) = a_0^{-j/2} \psi(a_0^{-j} t - k b_0). \tag{2.19}$$

The most commonly used dyadic scale adopts a geometric sequence with ratio, $r = 2$. Here, the values of the dilated and translated steps are set to $a_0 = 2$ and $b_0 = 1$, respectively. Using these two values, Equation 2.19 is simplified to

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j} t - k). \tag{2.20}$$

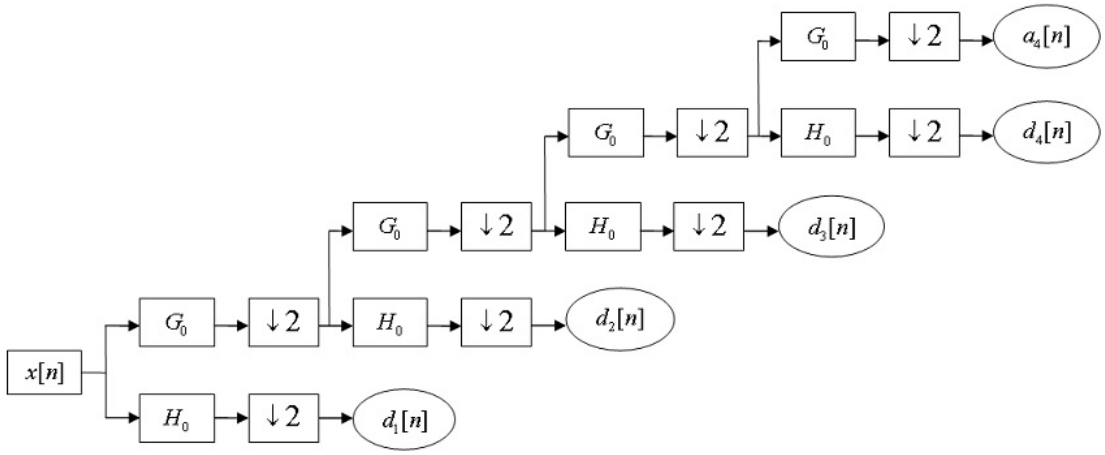

Figure 2.9: A four-level wavelet decomposition tree

DWT is computed using a series of low-pass and high-pass filters, known as decomposition trees, or filter banks, as shown in Figure 2.9. At each level, the original signal, $x(n)$ is decomposed into low and high frequencies using low-pass filters, G_0 and high-pass filters, H_0. Low-pass filters, G_0 yield approximation coefficients, $a(n)$, whereas high-pass filters, H_0 produce detail coefficients, $d(n)$. Each decomposition level reduces the time resolution into half of its original value. At the same time, the frequency resolution of the output signal is doubled. The decomposition process is repeated until the desired level of decomposition is achieved.

The fundamental tool used in calculating the wavelet coefficients of DWT is convolution, which is a binary operator. The convolution product, $\mathbf{y}$ of filter $\mathbf{h}$ and signal $\mathbf{x}$, denoted by $\mathbf{h} * \mathbf{x}$, is given by the following formula:

$$y_n = \sum_{k=-\infty}^{\infty} h_k x_{n-k}. \tag{2.21}$$

The filter $\mathbf{h}$ is basically a sequence of numbers that is used to pre-process a signal under study in order to obtain information embedded in the signal itself. The numbers are derived based on several conditions and assumptions. The derivation of the values of the Daubechies 4 (db4) filters (Van Fleet, 2011) is given in Appendix