ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

NEURO CLASSIFIERS FOR CONDITION AND BEARING HEALTH ASSESSMENT OF AN ELECTRIC MOTOR

M.Sc. THESIS

Mina GHORBAN ZADEH BADELI

Department of Electrical Engineering

Electrical Engineering Programme

JUNE 2022

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

NEURO CLASSIFIERS FOR CONDITION AND BEARING HEALTH ASSESSMENT OF AN ELECTRIC MOTOR

M.Sc. THESIS

Mina GHORBAN ZADEH BADELI
(504201032)

Department of Electrical Engineering

Electrical Engineering Programme

Thesis Advisor: Asst. Prof. Dr. Duygu BAYRAM KARA

JUNE 2022

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

ELEKTRİK MAKİNASINDA DURUM VE RULMAN SAĞLIĞI DEĞERLENDİRMESİ İÇİN NÖRO SINIFLANDIRICILAR

YÜKSEK LİSANS TEZİ

Mina GHORBAN ZADEH BADELI
(504201032)

Elektrik Mühendisliği Anabilim Dalı

Elektrik Mühendisliği Programı

Tez Danışmanı: Dr. Öğr. Üyesi Duygu BAYRAM KARA

HAZİRAN 2022

Mina GHORBAN ZADEH BADELI, a M.Sc. student of İTU Graduate School student ID 504201032, successfully defended the thesis/dissertation entitled "NEURO CLASSIFIERS FOR CONDITION AND BEARING HEALTH ASSESSMENT OF AN ELECTRIC MOTOR", which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Asst. Prof. Dr. Duygu BAYRAM KARA** İstanbul Technical University	....
Jury Members :	**Asst. Prof. Dr. Burak BARUTÇU** Istanbul Technical University	....
	Assoc. Prof. Dr. Ömer Cihan KIVANÇ Istanbul Okan University	....

Date of Submission : 3 June 2022
Date of Defense : 16 June 2022

To my family,

FOREWORD

First and foremost, I wish to sincerely thank my M.Sc. supervisor, Asst. Prof. Dr. Duygu Bayram Kara, who had constantly encouraged, challenged and treated me with respect. It is immeasurable how much I have learnt from her, and I am extremely grateful for working with her during my thesis study. It was my fortune and honor to have great supervision from Dr. Duygu Bayram Kara.

I would also like to thank the University of Tennessee Knoxville (UTK) for their assistance with the collection of the data set.

Last but by no means least, I would give my special gratitude to my family, especially my parents. They have always supported me spiritually throughout writing this thesis and encouraged me with their deepest love.

June 2022 Mina GHORBAN ZADEH BADELI

TABLE OF CONTENTS

ABBREVIATIONS

ADCNN	: Adaptive Deep Convolutional Neural Network
ANN	: Artificial Neural Network
BFGS	: Broyden-Fletcher-Goldfarb-Shanno
BP	: Backpropagation
CM	: Condition Monitoring
CNN	: Convolutional Neural Network
CWT	: Continuous Wavelet Transform
DB16	: Daubechies' 16
DFCNN	: Deep Fully Convolutional Neural Network
DFT	: Discrete Fourier Transform
DNN	: Deep Neural Network
DWT	: Discrete Wavelet Transform
EDM	: Electrical Discharge Machining
EM	: Electric motor
FD	: Fault Detection
FFT	: Fast Fourier Transform
FFNN	: Feedforward Neural Network
FN	: False Negative
FP	: False Positive
FPR	: False Positive Rate
HP	: High-pass filter
IM	: Induction motor
L-BFGS	: Limited-memory Broyden-Fletcher-Goldfarb-Shanno
LP	: Low-pass Filter
MCSA	: Motor Current Signature Analysis
MRA	: Multiresolution Analysis
MRWA	: Multiresolution Wavelet Analysis
MRWT	: Multiresolution Wavelet Transform
MSE	: Mean Squared Error
NN	: Neural Network

PMSM : Permanent Magnet Synchronous Motor
PRNN : Pattern Recognition Neural Network
PSD : Power Spectral Density
RELU : Rectified Linear Unit
RMS : Root Mean Squared
RNN : Recurrent Neural Network
ROC : Receiver Operating Characteristic
RPM : Rotation per minute
RUL : Remaining Useful Lifetime
STFT : Short-Time Fourier Transform
TN : True Negative
TP : True Positive
TPR : True Positive Rate
WT : Wavelet Transform

SYMBOLS

$\boldsymbol{x_i}$	: Input element (data samples) of an artificial neural network
$\boldsymbol{w_{ij}}$	: Interconnection weights
$\boldsymbol{b_j}$	: Bias value
$\boldsymbol{f}$	: Non-linear activation function
$\boldsymbol{Y_o}$	: Output of the neural network
$\boldsymbol{V_{jo}}$	: Interconnection weights between hidden layer and output layer
$\boldsymbol{J}$	: Jacobian matrix
$\boldsymbol{H}$	: Hessian matrix
$\boldsymbol{g}$	: Gradient of a function
$\boldsymbol{e}$	: Vector of errors for the network
$\boldsymbol{w_{k+1}}$	: Weight update at time $k+1$
$\boldsymbol{w_k}$	: Weight update at current time k
$\boldsymbol{\mu}$	: A scalar in update formula of the Levenberg-Marquardt algorithm
$\boldsymbol{\tilde{s}_k}$	: Second-order derivatives of a function
$\boldsymbol{\lambda_k}$	: Adjustable scalar parameter in Scaled Conjugate Gradient algorithm
$\boldsymbol{\tilde{p}_k}$	: Search direction vector in Scaled Conjugate Gradient algorithm
$\boldsymbol{\tilde{w}_k}$	: Weight vector at iteration k
$\boldsymbol{E(\tilde{w}_k)}$	: Global error function at iteration k
$\boldsymbol{E'(\tilde{w}_k)}$	: Gradient of error function at iteration k
$\boldsymbol{\alpha_k}$	: Scaled step size in Scaled Conjugate Gradient algorithm
$\boldsymbol{\tilde{\lambda}_k}$	: Updated value of λ_k
$\Delta_{\boldsymbol{k}}$	: Comparison parameter in Scaled Conjugate Gradient algorithm
$\boldsymbol{m}$	: Number of updates in L-BFGS algorithm
$\boldsymbol{\nabla f(x)}$	: Gradient of objective function at iteration k in L-BFGS algorithm
$\boldsymbol{H_k^0}$	: Initial approximate of inverse Hessian matrix in L-BFGS algorithm
$\boldsymbol{k}$	: Number of iterations
$\boldsymbol{d_k}$	: Approximate Newton's direction at iteration k
$\boldsymbol{H_k}$	: Inverse of the Hessian matrix at iteration k in L-BFGS algorithm
$\boldsymbol{g_k}$	: Gradient of objective function at iteration k in L-BFGS algorithm

$\boldsymbol{x_k}$	: Position at iteration k in L-BFGS algorithm
$\boldsymbol{H_{k+1}}$	: Inverse Hessian matrix update at iteration k
$\boldsymbol{x(t)}$	: Time-domain signal in Continuous Wavelet Transform
$\boldsymbol{\varphi}$	: Mother wavelet function
$\boldsymbol{a}$	: Scaling parameter in Continuous Wavelet Transform
$\boldsymbol{b}$	: Shifting parameter in Continuous Wavelet Transform
$\boldsymbol{W(a,t)}$	: Continuous Wavelet Transform
$\boldsymbol{DWT[j,k]}$	: Discrete Wavelet Transform
$\boldsymbol{\widetilde{H}}$	: Inverse low-pass filter used in MRWA
$\boldsymbol{\widetilde{G}}$	: Inverse high-pass filter used in MRWA
$\boldsymbol{D_n}$	: Detail coefficients in n-level MRWA
$\boldsymbol{A_n}$	: Approximation coefficients in n-level MRWA
$\boldsymbol{N}$	: Length of the signal
$\boldsymbol{H(x)}$	: Shannon Entropy for a signal
$\boldsymbol{P(x_i)}$	: Probability of obtaining the value x_i in Shannon Entropy
$\boldsymbol{P_x}$	: Power of a discrete-time signal $x[n]$
$\boldsymbol{P_{per\_power}}$	: Percentage of the total power in the 2-4 kHz frequency interval
$\boldsymbol{P_{2-4kHz}}$	: Average power of a signal in the 2-4 kHz frequency interval
$\boldsymbol{P_{total}}$	: Total power of a signal

LIST OF TABLES

Page

LIST OF FIGURES

Page

NEURO CLASSIFIERS FOR CONDITION AND BEARING HEALTH ASSESSMENT OF AN ELECTRIC MOTOR

SUMMARY

Electric motors (EMs) and their maintenance applications have recently constituted the central core of most industrial processes. Induction motors (IMs) are vastly used among various types of EMs due to their robustness, simplicity, and reliability. Indeed, reliability plays an essential role in all engineering applications. In order to enhance the integrity or safety and prevent any potential devastation or failure in the power systems, the Condition Monitoring (CM) of IMs has been encountered by lots of enthusiasts recently. Due to this, many researchers have focused on investigating and detecting the different types of failures in IMs. According to the information acquired from several surveys, the approximate rates of bearing failures, winding failures, and rotor-related failures are 40%, 30%, and 9%. It can be concluded that the bearing failures are the most probable faults in the IMs. Therefore, a CM technique is vital for detecting the bearing degradations at their earliest stages in an IM. In this way, accurate and quantitative information on the present condition of the IM can be found by tracking the performance of the motor.

There are plenty of CM and Fault detection (FD) techniques to diagnose the various faults and their associated frequencies in IMs. However, monitoring the aging levels indicating the overall condition of an IM is not considered vastly. Indeed, aging is a gradual and progressive process arising from several faults in the EMs. The bearing faults can be enumerated as the main problem accelerating the aging in an IM. Therefore, besides the overall condition of an IM, the bearing condition of the IM is also classified in this study. Furthermore, signals acquired from EMs contain lots of information on the condition of the system. The most favorable signal is the vibration since both the electrical and mechanical problems can be revealed through its proper interpretation. In this study, the vibration signals are employed due to their direct representations of the bearing degradations in an IM. Indeed, eight vibration signals have been collected through an accelerated aging experiment on an IM. The experimental data is obtained from an accelerated aging experiment, which was conducted in a research and development project supported by *The University of Tennessee Maintenance and Reliability Centre*. The Electrical Discharge Machining (EDM) and the thermal-chemical aging operations constitute the two phases of the accelerated aging experiment.

In recent years, Artificial Neural Networks (ANNs) have also been used vastly as simple and feasible computational tools for the CM and FD of IMs. There are many research works regarding the classification applications of ANNs in power systems. The preferable architecture for the classification applications of ANNs is the FFNN structure. The significant advantage of ANNs is the fact that they are able to classify various types of faults and their severities in an IM. In this study, two different types of Feedforward Neural Network (FFNN) models are designed with different training optimization algorithms for classifying eight aging levels of an IM. The first model is

a shallow Pattern Recognition Neural Network (PRNN) trained by two different optimization algorithms. The Levenberg-Marquardt and the Scaled Conjugate Gradient algorithms are employed as the two optimization algorithms in the developed PRNNs. The second classifier is a deep, fully connected FFNN solved by the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm. All the classification applications developed in this study benefit from the error Backpropagation (BP) technique.

The performances of these classifiers are primarily assessed with the aim of monitoring the overall condition of the IM. The second aim is to evaluate their performances in monitoring and classifying the different bearing conditions of the IM.

This study constitutes two parts containing four different ANN-based applications. In the first part, the time-domain statistical features of the provided experimental vibration signals are directly processed by different neuro classifiers to rate the overall conditions of the IM. Indeed, these two classifiers are operated without any preprocessing operations. This part evaluates the performance of the PRNN model trained by the Levenberg-Marquardt and the Scaled Conjugate Gradient algorithms and a multi-layered FFNN solved by the L-BFGS algorithm.

The second part represents the bearing aging classification applications by using the aforementioned shallow and deep classifiers. A comparison-based study is executed on the testing performances of the PRNN trained by the Levenberg-Marquardt algorithm and the L-BFGS-based FFNN. In this part, primarily, the high-frequency bands associated with the bearing faults in the spectra of the vibration signals are decomposed by using a one-level Multiresolution Wavelet Analysis (MRWA) technique. In this regard, only the essential frequency bands containing the bearing fault features are employed through a preprocessing process. Additionally, novel spectral-based features are selected as the inputs of the proposed classifiers to reduce the input dimensions of the network.

The performances of the networks, with the optimal hidden layer geometry in each application, are also analyzed by computing four metrics. The four metrics available for evaluating the developed classifiers are the accuracy rate, the Recall, the Precision, and the F1-score. In addition, the confusion matrix, Receiver Operating Characteristic (ROC), error histogram, and training loss plots are also used to assess the reliability and performance of the designed neuro classifiers. By comparing the performances of the proposed classifiers, it is deduced that the L-BFGS-based FFNN applications outperform the PRNN applications, generally. In the first part of the study, The testing accuracy rates of the network with the optimal geometry in the PRNN and the L-BFGS-based FFNN applications are 97.2% and 100%, respectively. Likewise, in the MRWA-based study, the testing accuracy rates of the network with the optimal geometry for the shallow and deep classifiers are 97.9% and 100%, respectively. Therefore, it is concluded that the outstanding performance of the MRWA-based deep classifier indicates the novelty and high reliability of the L-BFGS-based FFNN application in grading the different bearing aging levels of an IM. Indeed, among these four applications, MRWA-based FFNN solved by the L-BFGS algorithm has the most important contribution in terms of novelty, high reliability and high performance.

ELEKTRİK MAKİNASINDA DURUM VE RULMAN SAĞLIĞI DEĞERLENDİRMESİ İÇİN NÖRO SINIFLANDIRICILAR

ÖZET

Elektrik motorlarının durum izleme ve bakım uygulamaları son zamanlarda çoğu endüstriyel sürecin merkezini oluşturmuştur. Endüstride en çok kullanılan motor tipi olan Asenkron motorlar ise sağlamlıkları, basitlikleri ve güvenilirlikleri nedeniyle çeşitli motor türleri arasında yaygın olarak tercih edilmektedirler. Günümüz mühendislik uygulamalarında güvenilirlik çok önemli bir yere sahiptir. Güvenilirliği sağlamak ise sistem sürekliliğini, bütünlüğünü ve güvenliğini artırmak için güç sistemlerinde herhangi bir olası tahribatı veya arızayı önlemeyi gerektirmektedir. Bu sebeple asenkron makinelerin, durum izlemesi ve hata analizi konuları son zamanlarda çok ilgi çekmektedir. Bu nedenle, birçok araştırmacı asenkron makinalardaki farklı arıza türlerini araştırmaya ve tespit etmeye odaklanmıştır. Bu alanda yayınlanmış çalışmalara dayandırılmış derleme makale çalışmaları literatürde mevcuttur. Bu derleme makalelerin en önde geleninden elde edilen bilgilere göre; rulman arızaları, sargı arızaları ve rotorla ilgili arızaların yaklaşık oranları sırasıyla yaklaşık %40, %30 ve %9 şeklindedir. Diğer derleme makalelerin sonuçlarına göre de rulman arızalarının asenkron makinalarda en olası arızalar olduğu sonucuna varılmaktadır. Bu nedenle, bir asenkron makinada rulman bozulmalarını en erken aşamalarında tespit etmek için bir durum izleme tekniği hayati önem taşır. Bu sayede motorun performansı izlenerek makinanın mevcut durumu hakkında doğru ve nicel bilgiler bulunabilir.

Asenkron makinalarda çeşitli arızaları ve bunlarla ilişkili karakteristik frekansları teşhis etmek için çok sayıda durum izleme ve arıza tespit tekniği vardır. Bununla birlikte, bir makinanın genel durumunu gösteren yaşlanma seviyelerinin izlenmesi çok fazla dikkate alınmaz. Aslında yaşlanma, makinalardaki çeşitli hatalardan kaynaklanan kademeli ve ilerleyici bir süreçtir. Bir asenkron makinada yaşlanmaya katkıda bulunan temel hata modu olasılıksal olarak rulman arızaları olarak sayılabilir. Bu nedenle, bir asenkron makinanın genel durumunun yanı sıra, rulmanının sağlık durumu da bu çalışmada sınıflandırılmıştır.

Literatürdeki durum izleme ve hata belirleme çalışmaları model ve sinyal tabanlı olarak gerçekleştirilebilmektedir. Model tabanlı yaklaşımlar son derece pratik ve görece basit olmakla beraber genelleştirmeye uygun değillerdir. Bir başka deyişle model, ait olduğu sistemle ilgili doğru bilgiyi sağlayabilirken başka sistemlerde doğru çalışmayabilmektedir. Öte yandan sinyal tabanlı çalışmalarda tamamen sinyalin istatistiksel, spektral ve zaman – frekans özelliklerine dayalı yöntemler önerilmektedir. Bu yöntemler makina tipinden bağımsız olarak bir tutarlılık teşkil ederler. Bu anlamda sinyal tanablı durum izleme ve hata analizi yöntemleri daha genelleştirilebilir niteliktedir. Elektrik motorlarından durum izleme ve arıza tespiti amacıyla, akım, gerilim, titreşim, ses ve moment gibi farklı tiplerde sinyaller toplanabilmektedir. Bunlardan en yaygın olanı sinyal toplama kolaylığı göz önüne alınınca akım sinyalidir. Ancak akım sinyali elektriksel arızalar için direkt ve kolay indikasyonlar içerse de henüz büyümemiş sorunları mekanik sorunları çok da açık biçimde ortaya koymaz.

Sistemin durumu hakkında en genel bilgi titreşim sinyali ile elde edilir çünkü titreşim sinyali hem elektriksel hem de mekanik kısımlara ilişkin bilgileri bir arada sunma yetisine sahiptir.

Bu çalışmada da bir asenkron makinadaki rulman bozulmalarını doğrudan ortaya koyabilme yetileri nedeniyle titreşim sinyalleri kullanılmıştır. Deneysel veriler, A.B.D. Tennessee Üniversitesi Bakım ve Güvenilirlik Merkezi ile Oak Ridge Ulusal Laboratuarı tarafından desteklenen bir araştırma ve geliştirme projesinde yürütülen bir hızlandırılmış yaşlandırma deneyinden elde edilmiştir. Çalışmada asenkron makina üzerinde hızlandırılmış bir yaşlandırma deneyi yoluyla sekiz titreşim sinyali toplanmıştır. Bunlardan ilki sağlıklı durumda elde edilmiş ardından sistem yedi defa kombine bir yaşlandırma aksiyonuna maruz bırakılmıştır. Kombine yaşlandırma aksiyonu; elektriksel boşalma ve termokimyasal yaşlandırma işlemleri şeklinde iki adım içermektedir. Eletriksel boşalma adımı kapsamında mil üzerinde endüklenmesi muhtemel gerilimler ve bunların yaratacağı boşalmalar yapay olarak simüle edilmiştir. Termokimyasal yaşlandırma için ise motor belirli periodlarla su dolu bir tank içine batırılmış ve çıkarılıp endüstriyel fırınlarda yüksek ısıya maruz bırakılmıştır. Her yaşlandırma aksiyonu sonrasında motorun akım ve titreşim verileri farklı yük durumlarında kaydedilmiştir.

Son yıllarda, Yapay Sinir Ağları (YSA), asenkron makinaların durum izleme ve hata analizi için basit ve uygulanabilir hesaplama araçları olarak büyük ölçüde kullanılmıştır. YSA'ların güç sistemlerinde sınıflandırma uygulamaları ile ilgili birçok araştırma çalışması bulunmaktadır. Sınıflandırma uygulamaları için tercih edilen mimari ileri beslemeli tam bağlantılı yapay sinir ağı yapısıdır. Literatüre göre, YSA'ların önemli bir avantajı, asenkron makinalarda çeşitli arıza türlerini ve bunların önem derecelerini sınıflandırabilmeleridir. Bu çalışmada da bir asenkron makinanın sekiz yaşlandırma seviyesini sınıflandırmak için farklı eğitim optimizasyon algoritmaları ile iki farklı tip İleri Beslemeli Sinir Ağı (İBSA) modeli tasarlanmıştır. İlk model, iki farklı optimizasyon algoritması tarafından eğitilmiş sığ bir Örüntü Tanıma Sinir Ağıdır (ÖTSA). Geliştirilen ÖTSA'larda iki optimizasyon algoritması olarak Levenberg-Marquardt ve Scaled Conjugate Gradient algoritmaları kullanılmıştır. İkinci yapay sinir ağlı sınıflandırıcı, sınırlı bellekli Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algoritması tarafından çözülen derin, tam bağlantılı bir İBSA'dır. Bu çalışmada geliştirilen tüm sınıflandırma uygulamaları, hata geri yayılımı tekniğinden yararlanmaktadır.

Bu çalışma, dört farklı YSA tabanlı uygulamayı içeren iki bölümden oluşmaktadır. Bu sınıflandırıcıların performansları, öncelikle asenkron makinanın genel durumunu izlemek amacıyla değerlendirilir. İkinci amaç, asenkron makinanın farklı rulman koşullarını izleme ve durumlarını sınıflandırmaktır. Her iki bölüm için de ÖTSA ve L-BFGS ile eğitilen İBSA olmak üzere ikişer ağ yapılandırılmış ve optimize edilmiştir. Tüm bu dört uygulamada da optimum gizli katman geometrisine sahip ağların performansları dört metrik hesaplanarak analiz edilmiştir. Geliştirilen sınıflandırıcıları değerlendirmek için kullanılan dört metrik; doğruluk oranı, duyarlılık, kesinlik derecesi ve F1 puanıdır. Ayrıca, tasarlanan nöro-sınıflandırıcıların güvenilirliğini ve performansını değerlendirmek için karışıklık matrisi, karar vericinin etkinliği (ROC), hata histogramı ve eğitim kaybı grafikleri de kullanılmaktadır.

İlk bölümde, sağlanan deneysel titreşim sinyallerinin zaman domenli istatistiksel özellikleri, asenkron makinanın genel durumunu değerlendirmek için farklı nöro-sınıflandırıcılar tarafından doğrudan işlenmektedir. Bu konuda biri sığ (ÖTSA) diğeri derin (L-BFGS ile desteklenen İBSA) olmak üzere iki adet sınıflandırıcı yapılandırılmaktadır. Bu iki sınıflandırıcı herhangi bir ön proses işlemi olmaksızın ham verilerin ilgili istatistiksel parametreleri kullanılarak çalıştırılmaktadır. Bu bağlamda, Levenberg-Marquardt ve Scaled Conjugate Gradient algoritmaları tarafından eğitilen ÖTSA modelinin ve L-BFGS algoritması tarafından çözülen çok katmanlı bir İBSA'nın performansı değerlendirilmektedir. ÖTSA modelinin sonucu olarak Levenberg-Marquardt algoritması ile eğitilen ağların Scaled Conjugate Gradient algoritması ile eğitilen ağlardan daha iyi performans gösterdiği belirtilmektedir. Levenberg-Marquardt algoritması ile eğitilen en iyi yapının maksimum toplam doğruluk oranı %99,2 iken, bu ağ ile ilişkilendirilen test doğruluk oranı %97,2'dir. İlgili metrikleri analiz ederek ve farklı yapıların performans ölçüm grafiklerini karşılaştırarak, ÖTSA'nın optimum yapısının bile asenkron makinanın yaşlandırma sınıflandırması için yeterince güvenilir olmadığı sonucuna varılmıştır. Ayrıca optimum yapı için hesaplanan %100 eğitim doğruluk oranı, önerilen ağda aşırı uyum olasılığını temsil eder ki; bu YSA için istenen bir yapı değildir ve ağ ezberine işaret edebilir. Ezberleme riskini azaltmak ve daha güvenilir yapılar içeren bir model oluşturmak için, asenkron makinanın genel durumunu derecelendirmek amacıyla derin bir sınıflandırıcı olarak L-BFGS algoritması ile çözülen iki katmanlı bir İBSA modeli geliştirilmiştir. Bu sınıflandırıcının maksimum test doğruluk oranı %100'dür, bu da ÖTSA uygulamasından daha iyi bir performans gösterdiğini ortaya koymaktadır. Ayrıca, hesaplanan tüm metrikler ve performans ölçüm grafikleri, seçilen ağın kesinliğini ve doğruluğunu optimum bir yapı olarak onaylar. Bu uygulamada, önerilen derin sınıflandırıcı, derin YSA uygulamalarında fazla uyumu önleyerek, test veri setinin rastgele bölünmesi için güvenilir bir araç olan çapraz validasyon kullanmaktadır ve bu anlamda da ÖTSA uygulamasından daha güvenilir sayılmaktadır.

Çalışmanın ikinci kısmında yukarıda bahsedilen sığ ve derin sınıflandırıcılar kullanılarak rulman yaşı sınıflandırma uygulamaları yapılmıştır. Bu maksatla Levenberg-Marquardt algoritması tarafından eğitilen ÖTSA'nın ve L-BFGS tabanlı İBSA'nın test performansları üzerinde karşılaştırmaya dayalı bir çalışma yürütülmüştür. Bu bölümde öncelikle titreşim sinyallerinin spektrumlarındaki rulman arızaları ile ilişkili yüksek frekans bantları tek seviyeli Çok Çözünürlüklü Dalgacık Analizi (ÇÇDA) tekniği kullanılarak ayrıştırılmaktadır. Bu bağlamda, bir ön işleme süreci sayesinde yalnızca yatak arızası özelliklerini içeren temel frekans bantları kullanılmaktadır. Ek olarak, ağın girdi boyutlarını azaltmak için önerilen sınıflandırıcıların girdileri olarak yeni spektral tabanlı öznitelikler seçilmiştir. Hesaplanan metrikler ve performans ölçüm grafikleri incelenerek, önerilen ÖTSA'daki optimum yapının toplam ve test doğruluk oranlarının sırasıyla %95,6 ve %97,9 olduğu sonucuna varılmıştır. Optimum eğitim doğruluk oranı %100 olmadığı için bu uygulama daha güvenilir ve gerçekçidir. İlaveten veri setlerinin boyutu önceki uygulamaya kıyasla artırılarak ezberleme olasılığı azaltılmıştır.

Son olarak dördüncü sınıflandırıcı olarak, ÇÇDA tabanlı bir İBSA, L-BFGS algoritması tarafından, rulman yaşı sınıflandırma uygulamasında güvenilirlik ve performans açısından değerlendirilmek ve önerilen ÇÇDA-tabanlı ÖTSA ile karşılaştırılmak üzere eğitilir. Bu sınıflandırma aynı zamanda beş spektral temelli özniteliğe dayanmaktadır. Hesaplanan dört değerlendirme metriği ve performans ölçüm grafiklerinin analizi ile, minimum eğitim kaybı olan ağa ait maksimum test doğruluk oranının %100 olduğu tespit edilmiştir. Ayrıca, optimum sınıflandırıcı için performans; karışıklık matrisi, karar vericinin etkinliği ve eğitim kaybı grafikleri kullanılarak ortaya konmuştur. Hem sığ hem de derin sınıflandırıcılardaki optimum yapıların metriklerini, test karışıklık matrisini ve test karar vericinin etkinliği grafiklerini karşılaştırarak; L-BFGS kullanılarak rulman yaşı sınıflandırma uygulamasının performans ve güvenilirliğinin yüksek olduğu anlaşılabilir. Bu nedenle, ÇÇDA tabanlı derin sınıflandırıcı, rulman yaşlanma seviyesinin izlenmesi ve tespit edilmesi için en sağlam ve güvenilir model olarak önerilmektedir.

Önerilen sınıflandırıcıların hepsinin performansları karşılaştırılarak, L-BFGS tabanlı İBSA uygulamalarının genel olarak ÖTSA uygulamalarından daha iyi performans gösterdiği sonucuna varılmıştır. Çalışmanın ilk bölümünde, ÖTSA ve L-BFGS tabanlı İBSA uygulamalarında optimum geometriye sahip ağın test doğruluk oranları sırasıyla %97,2 ve %100'dür. Aynı şekilde ÇÇDA tabanlı çalışmada sığ ve derin sınıflandırıcılar için optimal geometriye sahip ağın test doğruluk oranları sırasıyla %97.9 ve %100'dür. Bu nedenle, ÇÇDA tabanlı derin sınıflandırıcının üstün performansı sayesinde bir asenkron makinanın farklı rulman eskime seviyelerini derecelendirmede yüksek güvenilirlik gösterdiği sonucuna varılmıştır. Nitekim bu dört uygulama arasında, L-BFGS algoritması ile çözülen ÇÇDA tabanlı İBSA, orjinallik, yüksek güvenilirlik ve yüksek performans açısından en önemli katkıya sahiptir.

Çalışmanın katkısı, spektral tabanlı öznitelikleri çıkararak ve yaşlanma izleme uygulamaları için verimli bir hibrit yapı sağlayarak yeni bir strateji ortaya koyması olarak özetlenebilmektedir. ÇÇDA tabanlı sınıflandırıcıların girdileri olarak dokuz istatistiksel öznitelik yerine beş spektral tabanlı öznitelik hesaplanarak, önerilen ağların boyutları da verimli bir şekilde azaltılmaktadır. Ayrıca, önerilen derin nöro-sınıflandırıcının yüksek performans ve doğruluk puanları, elektrik makinalarındaki rulman bozulmalarıyla ilişkili onarılamaz hasarları önlemek için büyük ölçüde durum izleme ve hata tespiti işlemlerini gerektiren endüstriyel uygulamalar için yardımcı olabilir.

1. INTRODUCTION

Electric motors (EMs) play an essential role in modern industry. In recent years, the most common types of EMs used in industrial applications were: DC motors, Induction motors (IMs), Permanent Magnet Synchronous Motors (PMSMs), switched reluctance motors, and brushless DC motors [1]. Among different types of EMs, IMs are the most widely used electro-mechanical devices in industrial applications due to their high reliability, simple structure, optimum efficiency, and less maintenance need [1,2]. Owing to these significant characteristics, Condition Monitoring (CM) and Fault Detection (FD) of IMs are vital issues that must be considered seriously [3]. Indeed, to prevent financial losses and physical damages, the health conditions of the IMs should be monitored and assessed at their earliest stages [4]. In addition, the quality of the performance and productivity of the system can be enhanced efficiently by investigating the health condition of the motor frequently and performing the early detection of inevitable faults [5].

Various electrical and mechanical faults occur in the IMs. Among all the electrical and mechanical faults, the bearing faults, which take place as mechanical faults, are the main reason for around 40% of the breakdowns in the IMs. Therefore, It is essential to note that by monitoring the health condition of IMs, different aging levels corresponding to the severity of these faults can also be detected accordingly. Some factors may interfere to form the bearing faults in the IMs; wear and tear, improper lubrication, overheating, shock loads, and imbalances are the other significant reasons that create the bearing faults in the IMs [6]. In recent years, many studies have focused specifically on the classification of the various faults in the IMs. However, the bearing aging classification application is rarely considered in the literature. Therefore, there are not enough research works that assess the performance of a classifier in terms of estimating different bearing degradation levels. That is why among all the classification-based studies, the bearing aging classification can be enumerated as a unique and favorable application.

1.1 Aim of the Thesis

The aim of this study is to classify the aging levels of an IM according to the overall condition of the motor and the bearing fault information, which is extracted from the experimental vibration data. From this point of view, the deduced aging class would be assessed as a condition estimation. Indeed, as a motivation, this study can lead us to analyze the performance of the motor precisely and provide proper life-extending precautions before encountering any costly losses. Thus, this study is beneficial for the CM and FD purposes, enhancing the quality of the system and preventing any malfunctioning or breakdown possibilities in the IMs.

This study is constituted of two parts determining the aging classification of the motor and rating the severity of the bearing degradation. In the first part, the age or condition classification will be implemented by using two different types of Neural Networks (NNs) based on statistical features. These two neural classifiers are developed to classify the vibration signals of an IM associated with different aging levels. The first one is a Pattern Recognition Neural Network (PRNN) trained by Levenberg-Marquardt and Scaled Conjugate Gradient algorithms. The second one is a fully connected Feedforward Neural Network (FFNN) solved by Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm. In the second part, for grading the bearing degradation, the Multiresolution Wavelet Analysis (MRWA) is used as the preprocessing operation in both shallow and deep classification applications. By employing the MRWA, the bearing fault-related part of the spectra is filtered out to distinguish the bearing fault features with clarity and enhance the performance of the proposed classifiers. In this section, new spectral features are introduced instead of statistical features to run the classifiers. The proposed hybrid structure can also be enumerated as a reliable method in terms of its novelty and efficiency in aging classification applications. Since most of the classification applications that have been developed recently were implemented to classify the different faults in the IMs and the aging classification is not appropriately investigated in the literature. Besides, using the L-BFGS algorithm as an optimization algorithm for a deeper neural classifier and employing the spectral-based features as the inputs of the network can also be a novel strategy for the aging classification purposes.

1.2 Literature Review

1.2.1 Aging problems of EMs and FD

Aging is a complex process involving unknown phenomena that can not be analyzed and estimated properly. In order to investigate the lifetime of the IMs or assess their aging levels, many researchers tried to concentrate specifically on the aging-related subjects despite the limitations and problems associated with the aging process [7-9]. Besides, compared to electrical signals, mechanical signals such as vibration signals will be the better indicators of aging since the vibration signals provide the aging information related to both electrical and mechanical parts [10]. Indeed, the vibration signal contains a stochastic nature due to the existence of some noises combined with the frequency components in it. Because of this fact, analysis or detection of the aging with these signals would be a challenging task.

Vibration signals can be generated due to the different mechanical faults in the IMs. The bearing fault is actually the main contributor to the vibration oscillations and aging process in an IM [11-13]. The bearing faults are prevalent and inevitable for EMs, especially for IMs that have been dominating the industry in recent years [6]. Indeed, between 40% and 50% of the faults that occur in the IMs are related to bearing defects. Therefore, bearing faults have the highest rate of breakdowns in the IMs. There are three types of mechanical defects in the IMs that can be classified as ball-bearing faults. These defects are categorized as ball defects (cracking and wearing), inner bearing race defects, and outer bearing race defects (wearing or corrosion) [2].

As it is mentioned before, the aging levels of an IM can be determined simply by analysis of the vibration signal and monitoring the bearing health. Indeed, there are two general approaches for the monitoring and age determination of the EMs. The first method is to estimate the Remaining Useful Lifetime (RUL) of the system. This method is very difficult to be implemented. Since so many datasets must be provided and analyzed in this process. The second approach is to deduce a general class determination broadly on the age of the system. This method can be preferable, and it can also be managed by proper accelerated aging experimental data.

Thus, for the CM and FD of IMs, some useful signal-based techniques have been developed to detect failures and prevent sudden breakdowns and probable damages in the IMs. The Motor Current Signature Analysis (MCSA) is one of the most common methods that use the Fast Fourier Transform (FFT) and Power Spectral Density (PSD) of the signals to diagnose the faults in an IM [14]. The Short-Time Fourier Transform (STFT) method is also used to analyze the signals in both time and frequency domains [15]. However, the STFT method has some shortcomings, such as poor frequency resolution due to the constant window size that it uses for all frequencies [16]. Therefore, in order to overcome these problems, the researchers started to replace the STFT method with the Wavelet Transforms (WT) for CM and FD purposes [17,18]. There are also many studies in which the Discrete Wavelet Transform (DWT) was preferred for the CM and FD goals [19,20]. Among all the presented classical methods, Multiresolution Analysis (MRA) plays an essential role in the CM and bearing FD process. MRA can be treated as a digital filter by its nature. Hence, the MRA can be an appropriate way to detect bearing faults by using the vibration signals in different frequency bands with different resolutions [21].

On the other hand, many novel CM and FD methods have emerged that outperform the conventional signal-based methods. Indeed, one of the modern and precious techniques for bearing fault classification purposes in the IMs is the application of Artificial Neural Networks (ANNs) [22]. In the literature, various types of faults are defined for the IMs that progress the aging [2,4]. In this regard, a classifier labeling the condition broadly as a rate instead of detecting and naming the faults might be more feasible and useful for CM purposes.

1.2.2 Neural Classifiers for CM and FD

It is important to note that different types of signals may be used to monitor the condition of the IM, such as voltages [23], currents [24], vibrations, and temperatures [25]. Nevertheless, vibration monitoring is the most informative CM technique since it contains both mechanical and electrical fault indications. Hence, most applications associated with bearing failure detection are based on vibration signals [26].

Fortunately, with the improvement of automated tools, conventional signal processing methods have been replaced with soft computing techniques over time. ANNs are one of the well-known methods of soft computing techniques that have been applicable for FD of power systems in recent years. There are different types of vibration-based ANN applications for the CM and fault identification of the IMs. Predominantly, these differences are related to the type of features that can be extracted from the vibration signals and the type of the developed NNs that will be availed as the classifiers. Some researchers prefer to use only the statistical parameters for the fault classification applications in the IMs, which are quite handy and prevalent [27,28]. On the other hand, many FD-based studies are successful with the time-frequency-domain features [28-30]. Nevertheless, in some studies, it is approved that the efficiency of the time-domain statistical features extracted from the vibration signals is better than the time-frequency-domain features used in fault classification applications [28]. Indeed, the time-domain analysis represents the statistical characteristics of the vibration signal, such as peak level, standard deviation, skewness, kurtosis, and crest factor. However, a frequency-domain analysis is based on the different interpretations of the Fourier transform, such as amplitude or power [31]. The vibration power and different fault frequency amplitudes calculated with FFT-based methods are the parameters that have been used frequently as the frequency-domain features in the literature [30,32]. Likewise, since the vibration signals are non-stationary signals, conventional signal processing methods such as FFT-based methods are not suitable enough to manifest the inherent information of the non-stationary signals [31]. Therefore, it is evident that there is a limited number of studies using just spectral features for the FD aims in the IMs.

Through the assessment of the hybrid classification applications, it can be deduced that the combination of the wavelet-based methods and the ANNs has enhanced the reliability of the bearing fault classifications in the IMs [33]. The wavelet-based methods are very beneficial in terms of extracting the specific frequency bands in the vibration signals. Indeed, wavelet transform is widely employed in the FD and CM of IMs. However, for the bearing fault classification aims, a limited number of studies investigate the efficiencies of the different types of classifiers that use the wavelet-based features as their input parameters [30].

As mentioned before, the type of the ANN used as the classifier is the significant factor that distinguishes the bearing fault classification applications from each other. Since each type of NN represents a specific performance for the same vibration data set. Many researchers have evaluated the performances of both shallow NNs and Deep Neural Networks (DNNs) as the bearing FD applications in the IMs [34-37]. The FFNN trained with the Backpropagation (BP) algorithm is principally used to detect the bearing faults in the IMs [27-32,34,35].

There are many substantial advantages to the usage of DNNs instead of shallow NNs in various applications. One of the important reasons that DNNs are preferred mostly can be associated with the fact that their training algorithms are evolved in an appropriate way to control the training process with a faster speed, better convergence, and improved generalization. Applications of the deeper NNs for bearing FD purposes in an IM have been presented as eligible and popular case studies recently. As an example, the first application of a Convolutional Neural Network (CNN) for the bearing FD aim in the IMs was implemented in 2016 [36]. During the next three years, many researchers attempted to apply the same technique for the CM and fault identification of the electric machines [38-52]. Therefore, many advances have emerged in bearing FD techniques based on DNN algorithms. The Adaptive Deep Convolutional Neural Network (ADCNN) [38] and the Deep Fully Convolutional Neural Network (DFCNN) [43] are suitable examples of the advanced DNNs used for the bearing fault classification aims. Other types of DNNs, such as auto-encoders, have been employed numerously for the FD and CM of electric machines. In some studies, the accuracy rates of the developed auto-encoder-based networks could successfully reach 99.6% [53] and 99.83% [54]. Recurrent Neural Networks (RNNs) are also applicable as DNNs for the classification of the bearing health conditions in electric machines [55,56]. It is obvious that the scope of the studies assessing the performance of the different types of DNNs for the classification of the bearing faults in an IM is very vast.

However, for the bearing aging classification aim, there is not any reliable study containing the L-BFGS-based classifier that is used both the MRWA and spectral-based features as its input parameters. Indeed, the main reason for using a deep neural classifier is the fact that the DNNs can operate as a shortcut to understanding the age of the motor in terms of the bearing health condition.

1.3 Hypothesis

In this thesis, the significant contributions are focused on the efficiency and higher reliability of the L-BFGS-based FFNN as a deep network compared to the PRNN. Indeed, DNN might be a better solution for the NN-based bearing fault classification applications. Furthermore, it is an undeniable fact that using the statistical parameters is still a very handy way to label the motors in terms of their various aging or condition levels. Owing to this fact, in most of the NN-based CM and FD applications, statistical parameters or the combination of the statistical parameters and the frequency-domain parameters are still utilized vastly. In this study, with the aim of improving the robustness of the proposed shallow and deep NNs, two methods are also considered to be executed simultaneously. The first one is to apply the MRWA to the acquired vibration signals. Since, by applying the MRWA, a specific frequency band might be used as a better indication for the bearing degradation. Besides, as the second preprocessing step, new spectral-based features can be presented for band-specific NN classification applications. In this way, the dimension of the inputs of the proposed networks can also be reduced efficiently.

2. MATHEMATICAL BACKGROUND

In this section, all the mathematical methods used for the aging classification of the IMs are explained. Primarily, the mathematical background of the ANN is discussed in detail. Then, the Levenberg-Marquardt, the Scaled Conjugate Gradient, and the L-BFGS algorithms are introduced as the optimization algorithms of the proposed bearing aging classifiers. Eventually, the mathematical background of MRWA is also represented with clarity.

2.1 Artificial Neural Networks (ANNs)

ANNs can be defined as very functional tools acting very similarly to biological neurons. In fact, ANNs consist of straightforward, distributed, and strong processing units called artificial neurons, which can be used to emulate the behavior and the functions of the biological neurons existing in the human brain. These simple elements of an ANN are able to support intelligent data evaluation techniques, such as pattern recognition, classification, and generalization [57]. Three significant concepts must be defined before designing a simple ANN model are:

- Topology of the network
- Type of the training algorithm (learning procedure)
- Type of the activation functions

Among the different types of NNs, the multi-layered FFNN is the first and the simplest one [58]. There are three layers of nodes in the structure of a multi-layered perceptron network: an input layer, one or more hidden layers, and an output layer. Neurons belonging to a layer have to be connected to neurons of a different layer [59]. Besides, the information in this network flows from the input layer directly, through any hidden layers, to the output layer without any cycle or feedback [58]. A three-layered FFNN model is demonstrated in Figure 2.1. In the multi-layered perceptron NNs, each input element, $x_i(i = 1,2, \dots, n)$, is multiplied by an interconnection weight, w_{ij}, and then the weighted inputs are inserted into a sum function for each neuron. Afterward, by the sum of a bias value, b_j, and the weighted sum of the input data for each neuron, the resultant P_j can be calculated. The mathematical model of the resultant input, P_j, for neuron j is presented in equation (2.1).

$$P_j = \sum_{i=1}^{n} w_{ij} x_i + b_j \tag{2.1}$$

Afterward, a non-linear activation function, f, is applied to the resultant input, P_j, in the hidden layer. The mathematical model for the outputs of the network, $Y_o(o = 1,2,\ldots,k)$, is represented in equation (2.2).

$$Y_o = \sum_{j=1}^{m} V_{jo}\, f(P_j) \tag{2.2}$$

where V_{jo} is defined as the interconnection weight of the nodes between the hidden layer and the output layer.

Assigning the proper training algorithm is a significant issue that affects the success of prediction in the NNs. Many training methods have been established to optimize the learning procedure of the ANNs. The most prevalent technique is the BP algorithm which adjusts the weights and biases by using a backward pass after each forward pass through the network. Indeed, in the BP algorithm, the difference between output and the actual value is computed as the error. Afterward, the resultant error is fed back towards the input layer through the network. By using the produced information, the algorithm adjusts the weights of each connection repeatedly to minimize the error function [60]. Levenberg–Marquardt, Bayesian Regulation, and Scaled Conjugate Gradient are some of the useful BP-based training algorithms used for ANN applications. After completing the training process, the NN is usually tested with the new data sets.

In the first part of the study, both Levenberg–Marquardt and Scaled Conjugate Gradient BP algorithms are examined as the optimization algorithms of the developed PRNN, which analyzes the statistical features. Besides, the L-BFGS-based FFNN is also developed for the classification of the statistical-based inputs. However, In the second part of the study, a comparison-based work is represented just for the FFNN trained with the Levenberg–Marquardt algorithm and the deep FFNN trained with the

L-BFGS algorithm. Further information regarding the topologies of the developed ANNs is presented in detail in Sections 4 and 5, respectively.

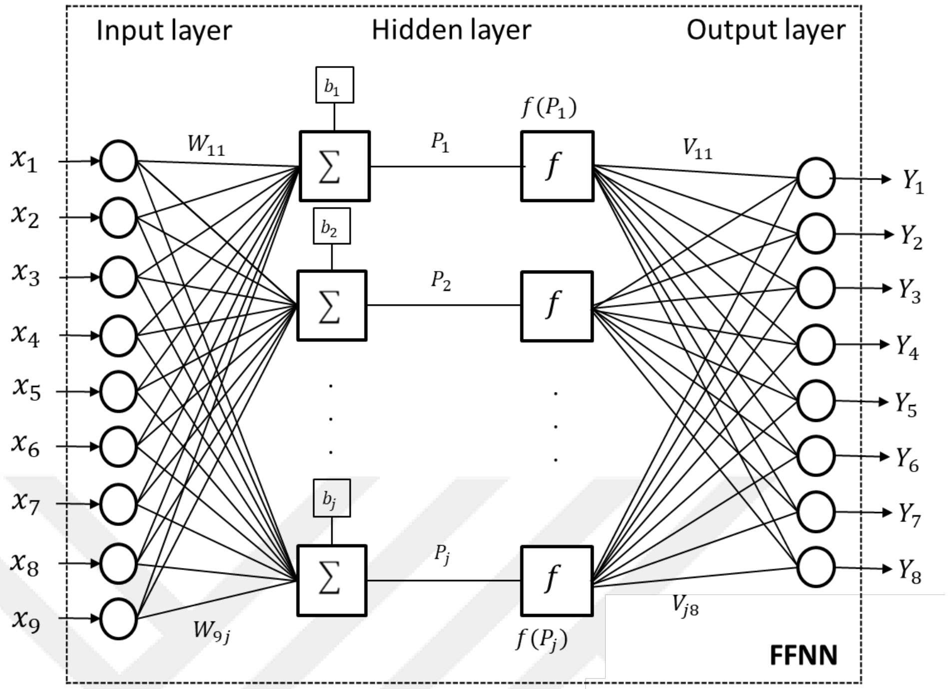

Figure 2.1: A three-layered FFNN architecture.

2.1.1 PRNN trained by Levenberg-Marquardt and Scaled Conjugate Gradient

The PRNN is a type of FFNN developed to be applicable for shallow NN-based classification demands. In the PRNNs, the target values are defined just by vectors of zero values and a 1, which represents the desired specific class. The Levenberg-Marquardt and the Scaled Conjugate Gradient algorithms are usually availed as the training functions in the PRNN.

The Levenberg-Marquardt algorithm [61] contains two significant attributes affecting the performance of the NNs in terms of their speed and convergence. Indeed, The Levenberg-Marquardt algorithm operates as a hybrid method in which the useful features of the steepest descent method and the Gauss-Newton algorithm are combined. In other words, the Levenberg-Marquardt algorithm benefits from the high speed of the Gauss-Newton algorithm and the stable convergence of the steepest descent method. The Levenberg-Marquardt algorithm is created to solve the problem regarding the minimization of the performance functions, which comprise the sum of the squared errors [62]. This algorithm solves the minimization problem by using a Jacobian matrix, J, for the approximation of the Hessian matrix, H, which is defined in equation (2.3). Therefore, the second-order training speed can be obtained without any difficulties in calculating the Hessian matrix.

$$H = J^T J \tag{2.3}$$

In the Levenberg-Marquardt algorithm, the Jacobian matrix is also determined as the first derivative of the network errors with respect to the weights and biases. Indeed, the computational complexity can be reduced by utilizing the Jacobian matrix instead of the Hessian matrix. Because for the calculation of the Jacobian matrix, a technique based on the standard BP algorithm is used [61]. Moreover, the calculation of the gradient, g, can also be performed according to equation (2.4),

$$g = J^T e \tag{2.4}$$

where e is defined as a vector of errors for the network.

The update formula used in the Levenberg-Marquardt algorithm is very similar to Newton's update formula. Equation (2.5) represents the update formula of the Levenberg-Marquardt algorithm containing the approximation of the Hessian matrix,

$$w_{k+1} = w_k - [J^T J + \mu I]^{-1} J^T e \tag{2.5}$$

where w_{k+1} is the weight update at time $k+1$ and w_k is the weight update at current time k. If the scalar μ is equal to zero, the formula operates as Newton's method in which the approximated Hessian matrix is availed. By using an increased value of μ, the formula operates as the gradient descent method with a small step size. It is preferred to use the decreased value of μ to reach a higher speed and better accuracy, which occurs near the error minimum in Newton's method.

The Scaled Conjugate Gradient algorithm [63] is also designed to operate as an automated supervised optimization algorithm without having to do the line search in each learning iteration and depending on the parameters which are selected by the user. Indeed, the line search in each iteration is enumerated as a time-consuming process that can be eliminated by such an automated algorithm. On the other hand, choosing the best value for the user-dependent parameters in the optimization algorithms is a challenging step affecting the efficiencies of the networks. As an example, the learning rate is a user-dependent parameter that affects the performance of the standard BP algorithm. Indeed, the learning rate assigns the value of the step size at each iteration in an optimization algorithm intending to reach the minimum of the loss function. In order to find an optimum value for the step size, the line search per learning iteration is also implemented in other optimization algorithms such as Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. Somehow the Scaled Conjugate Gradient algorithm omits these difficulties by estimating the step size using a different technique. Basically, the Scaled Conjugate Gradient algorithm is the improved version of the Conjugate Gradient algorithm combined with the Levenberg-Marquardt approach to omit the line search process in each iteration and scale the step size. In order to summarize the mathematical background of the Scaled Conjugate Gradient algorithm, a new mathematical representation for the second-order term, $\tilde{s}_k = E''(\widetilde{w}_k)\tilde{p}_k$, is introduced in equation (2.6) primarily.

$$\tilde{s}_k = \frac{E'(\widetilde{w}_k + \sigma_k \tilde{p}_k) - E'(\widetilde{w}_k)}{\sigma_k} + \lambda_k \tilde{p}_k \tag{2.6}$$

where λ_k is an adjustable scalar parameter that is regulated according to the sign of δ_k in each iteration. $\tilde{p}_k$ is denoted as the search direction vector, and $\widetilde{w}_k$ is the weight vector at iteration k. Moreover, in the Scaled Conjugate Gradient algorithm, $E(\widetilde{w}_k)$ and $E'(\widetilde{w}_k)$ are defined as the global error function and the gradient of error at iteration k, respectively. The value of σ_k can also be calculated according to the formula represented in equation (2.7), where the initial condition is also set as $0 < \sigma \leq 10^{-4}$.

$$\sigma_k = \frac{\sigma}{|\tilde{p}_k|} \tag{2.7}$$

Equation (2.8) indicates the scaled step size formula developed in the Scaled Conjugate Gradient algorithm.

$$\alpha_k = \frac{\mu_k}{\delta_k} = \frac{-\tilde{p}_k{}^T E'(\widetilde{w}_k)}{\tilde{p}_k{}^T \tilde{s}_k + \lambda_k |\tilde{p}_k|^2} \tag{2.8}$$

λ_k can also be updated as $\tilde{\lambda}_k$ according to the formula demonstrated in equation (2.9).

$$\tilde{\lambda}_k = 2(\lambda_k - \frac{\delta_k}{|\tilde{p}_k|^2}) \tag{2.9}$$

If $\Delta_k > 0.75$, then $\lambda_k = \frac{1}{4}\lambda_k$.

If $\Delta_k < 0.25$, then $\lambda_k = \lambda_k + \frac{\delta_k(1-\Delta_k)}{|\tilde{p}_k|^2}$.

where Δ_k is defined as a comparison parameter which is formulated in equation (2.10).

$$\Delta_k = \frac{2\delta_k[E(\widetilde{w}_k) - E(\widetilde{w}_k + \alpha_k \tilde{p}_k)]}{{\mu_k}^2} \tag{2.10}$$

The initial values of λ_l and $\tilde{\lambda}_l$ are assigned to be as $0 < \lambda_l \leq 10^{-6}$ and $\tilde{\lambda}_l = 0$, respectively.

2.1.2 FFNN trained by L-BFGS

In this thesis, a fully connected multi-layered FFNN training with the L-BFGS algorithm is developed as a classification model. Since the L-BFGS algorithm requires less memory than the other large-scale optimization algorithms, it has been listed as a suitable algorithm for many DNN applications recently. Indeed, in the L-BFGS algorithm, the approximated inverse Hessian of the objective function $f(x)$ is constructed by using the history of gradient evaluations. Moreover, to train a DNN by the L-BFGS algorithm, the internal parameters of the DNN must be tuned by the L-BFGS algorithm. In this way, the energy of the error signal between the training data set and the output of the DNN is minimized efficiently. The L-BFGS algorithm was introduced for the first time in 1980 [64]. The L-BFGS algorithm is very similar to the BFGS algorithm. However, there is a significant difference between these two algorithms in the methods they have to use to update the Hessian matrix. In other words, a full history of the differences of the gradients over several iterations will be availed to approximate the Hessian matrix by the BFGS algorithm, while the L-BFGS algorithm is based only on the most recent m gradients. Obviously, the storage requirement corresponding to m gradients is much smaller than the storage required for the full history of the gradients.

The L-BFGS algorithm can be categorized as an optimization algorithm in the family of quasi-Newton methods [65]. The L-BFGS algorithm approximates the Hessian matrix (second derivative) of $f(x)$ by using a reduced amount of memory. Indeed, unlike the BFGS algorithm, the L-BFGS avails a small amount of the updates (m) of the position x and gradient $\nabla f(x)$, and it does not operate with the knowledge of the sparsity structure of the Hessian. In other words, in the L-BFGS algorithm, after selecting the m corrections of the BFGS algorithm by the user, the initial approximate of the inverse Hessian matrix H_k^0 is selected to start the estimation process at iteration k. The L-BFGS algorithm works the same as the BFGS algorithm during the first m iterations. However, for $k > m$, the approximation process of the inverse Hessian matrix is different in the L-BFGS algorithm. Indeed, for $k > m$, the information of the m updates of the BFGS algorithm is applied to the H_k^0, and then the updated value of H_k will be acquired efficiently.

Therefore, it can be an appropriate method for optimization applications containing high-dimensional minimization problems. Moreover, their linear memory requirement corroborates their simplicity and high speed in computational operations. In the L-BFGS algorithm, the differentiable scalar function $f(x)$ can be minimized over unconstrained values of the real-vector x efficiently.

In order to describe the L-BFGS algorithm in a superior way, it is required to introduce the mathematical notations of the proposed method. In the first step, the algorithm needs an initial estimate of the optimal value, x_0, to be improved by using a sequence of better estimates, such as $x_1, x_2, \ldots$, iteratively. On the other hand, the gradient of the function that is going to be minimized can be defined as $g_k = \nabla f(x)$. Indeed, by using the gradient of the function, the direction of the steepest descent is determined, and also the approximation of the Hessian matrix can be acquired effectively. Let d_k be the approximate Newton's direction and H_k be the inverse of the Hessian matrix, and then g_k is the current gradient. As seen in equation (2.11), by multiplication of the H_k and g_k, the approximate Newton's direction is computed implicitly. Actually, there is a significant difference between the L-BFGS algorithm and other quasi-Newton's algorithms in performing the mentioned matrix-vector multiplication.

$$d_k = -H_k g_k \tag{2.11}$$

In order to construct the direction vector, a prevalent approach that is known as the "two-loop recursion" is availed in the L-BFGS algorithm. Indeed, this technique is performed by using the history of the updates [64].

Primarily, we define $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$ where x_k is assumed to be the position at the k_{th} iteration. In addition, it is assumed that the last m updates of the form $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$ are stored properly.

For computing the inverse Hessian matrix, the proposed algorithm uses the BFGS-based update formula, which is demonstrated in product form in equation (2.12),

$$H_{k+1} = (I - \rho_k s_k y_k^T) H_k (I - \rho_k y_k s_k^T) + \rho_k s_k s_k^T \tag{2.12}$$

where $\rho_k = \frac{1}{y_k^T s_k}$ and $y_k^T s_k > 0$ for all values of k.

The H_k^0 is also defined as the initial approximate of the inverse Hessian that the estimation at iteration k begins with. In order to have a numerically appropriate estimation for H_k^0, it is preferred to select a positive definite and diagonal matrix or a multiple of the identity matrix as the initial approximate of the inverse Hessian matrix. Indeed, according to equations (2.13) and (2.14), H_k^0 can be scaled in a reliable way to ensure that the unit step length is accepted in most iterations.

$$\gamma_k = \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}} \tag{2.13}$$

$$H_k^0 = \gamma_k I \tag{2.14}$$

Furthermore, the validity of the curvature condition and the stability of the BFGS updating can be confirmed by using the Wolfe line search. It is required to select a proper step length to satisfy the Wolfe conditions [66,67] that are introduced in equations (2.15) and (2.16). Indeed, equation 8 and equation 9 are known as Armijo rule and curvature condition, respectively.

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \beta' \alpha_k d_k^T g_k \tag{2.15}$$

$$d_k^T g(x_k + \alpha_k d_k) \geq \beta d_k^T g_k \tag{2.16}$$

In point of fact, we choose x_0, m, $0 < \beta' < \frac{1}{2}$, $\beta' < \beta < 1$, and a symmetric and positive definite initial Hessian matrix H_k^0 at the first step of the L-BFGS algorithm. Nocedal, J. [64] has also determined the proper example values of β' and β for the quasi-Newton methods. In the proposed research work, β' and β are selected to be equal to 10^{-4} and 0.9, respectively [64].

Likewise, a recursive formula has been assigned for computing the multiplication of the inverse Hessian and the gradient of the objective function [64]. In the proposed formula, primarily for a fixed k, we define a sequence of vectors $q_{k-m}, \dots, q_k$ as $q_k = g_k$ and $q_i = (I - \rho_i y_i s_i^T) q_{i+1}$. Then a recursive algorithm for calculating q_i from q_{i+1} is to define $\alpha_i = \rho_i s_i^T q_{i+1}$ and $q_i = q_{i+1} - \alpha_i y_i$. We also define another sequence of vectors $d_{k-m}, \dots, d_k$ as $d_i = H_i q_i$. There is another recursive algorithm for calculating these vectors which is to define $d_{k-m} = H_k^0 q_{k-m}$ and then recursively define $\beta_i = \rho_i y_i^T d_i$ and $d_{i+1} = d_i + (\alpha_i - \beta_i) s_i$. The value of d_k is then the ascent direction. Therefore, the descent direction can be ascertained according to the algorithm demonstrated in Figure 2.2. It is important to note that by using the formula which was introduced in equation (2.11), the search direction can be determined for the minimization process precisely.

ALGORITHM:

- $\boldsymbol{q = g_k}$
- For $\boldsymbol{i = k-1, k-2, \dots, k-m}$
  - $\boldsymbol{\alpha_i = \rho_i s_i^T q}$
  - $\boldsymbol{q = q - \alpha_i y_i}$
- $\boldsymbol{\gamma_k = \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}}$
- $\boldsymbol{H_k^0 = \gamma_k I}$
- $\boldsymbol{d = H_k^0 q}$
- For $\boldsymbol{i = k-m, k-m+1, \dots, k-1}$
  - $\boldsymbol{\beta_i = \rho_i y_i^T d}$
  - $\boldsymbol{d = d + s_i(\alpha_i - \beta_i)}$
- $\boldsymbol{d = -d}$

Figure 2.2: The recursive method for computing the descent direction in L-BFGS algorithm.

2.2 Multi Resolution Wavelet Analysis (MRWA)

The MRWA, which is based on the DWT, was first presented by S. G. Mallat [68]. The MRWA is a prominent tool for decomposing the signals into different levels of resolution. In other words, the coarse resolution and the finer resolution of the signal exhibit the information about the higher-frequency components and lower-frequency components of the signal, respectively. In order to create a series of high and low pass filters that divide the spectrum into two parts, a scaling function or low-pass filter (LP) and a wavelet function or high-pass filter (HP) are assigned in the Multiresolution Wavelet Transform (MRWT) formula. This type of filtering is implemented iteratively for each level of decomposition and reconstruction.

Indeed, the whole process of MRWA consists of two general steps. The first step belongs to the decomposition of the signal and the second step is for the reconstruction process [69]. In order to understand the concept of the MRWA, it is necessary to introduce the mathematical expression of the WT primarily. WTs are used for the time-frequency analysis of a signal generally [70]. Indeed, the sum of multiplications of the original signal with the scaled and shifted wavelets constitutes the WT. In equation (2.17), the mathematical definition of the Continuous Wavelet Transform (CWT) for a signal $x(t)$ is demonstrated explicitly [69].

$$W(a,t) = \frac{1}{\sqrt{a}} \int x(t) \varphi^{*}(\frac{t-b}{a})\, dt \tag{2.17}$$

In equation (2.17), function φ is defined as the mother wavelet. The mother wavelet is a kernel function having a special waveform with finite energy and defined in a finite interval. It is important to note that the selection of the mother wavelet function should be accomplished by considering the signal to be observed. Moreover, in equation (2.17), a and b are defined as continuous scaling and shifting parameters, respectively. These parameters can be discretized to form the DWT of a signal [69,70]. The DWT helps us to represent the signal by the approximation coefficients and detail coefficients indicating the low-frequency and high-frequency components, respectively [19]. The mathematical definition of the DWT for a signal $x[n]$ is ascertained in equation (2.18).

$$DWT[j,k] = \frac{1}{\sqrt{{a_0}^j}} \sum_{n} x[n] \varphi \left[\frac{k - n b_0 {a_0}^j}{{a_0}^j} \right] \tag{2.18}$$

where scaling coefficient $a = {a_0}^j$ and shifting coefficient $b = n b_0 {a_0}^j$. In addition, it is important to mention that $n, j \in Z$, $a_0 > 1$, and $b_0 > 0$.

In MRWA, for each level of decomposition, a LP filter H and a HP filter G are applied to the signal, and then the downsampling by two is executed on the outputs of the wavelet filters. Afterward, the inverse transform is availed for the reconstruction process. In this step, before applying the inverse low-pass filter $\widetilde{H}$ and the inverse high-pass filter $\tilde{G}$, an upsampling by two is performed for each level of reconstruction. As an example, Figure 2.3 demonstrates the one-level decomposition and reconstruction process of the MRWA. In order to prevent information redundancy, decimators and interpolators are used in the up or downsampling process [71].

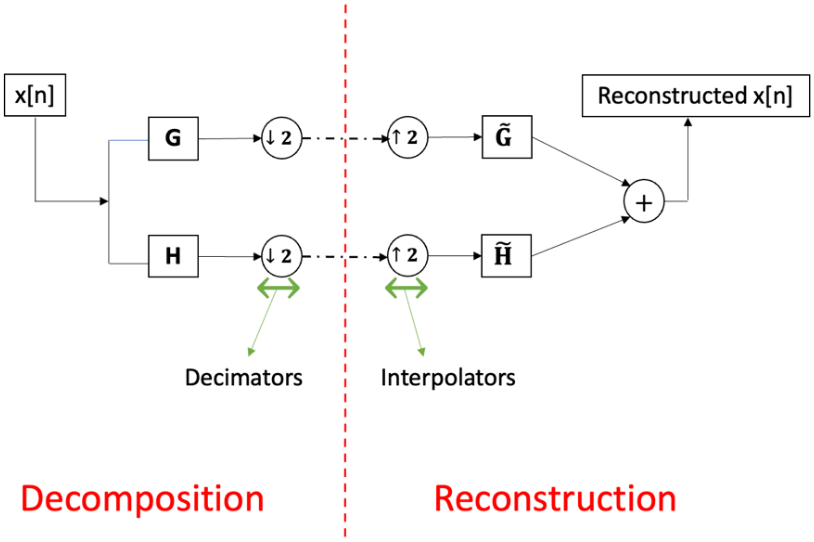

Figure 2.3 : Schematic interpretation of the decomposition and reconstruction process of a one-level MRWA.

In order to attain a higher resolution or reach a specific frequency region of the signal, more than one level of decomposition can be specified for the MRWA. According to Figure 2.4, in which the structure of the n-level MRWA is illustrated, the outputs of the high and low-frequency filters are defined as details D_n and approximations A_n, respectively. Eventually, the signal can be reconstructed by the sum of the n_{th} approximation coefficient and all the acquired details from each level. Equation (2.19) represents the reconstruction formula for the signal $x(t)$ [69].

$$x(t) = D_1 + D_2 + D_3 + \cdots + D_n + A_n \tag{2.19}$$

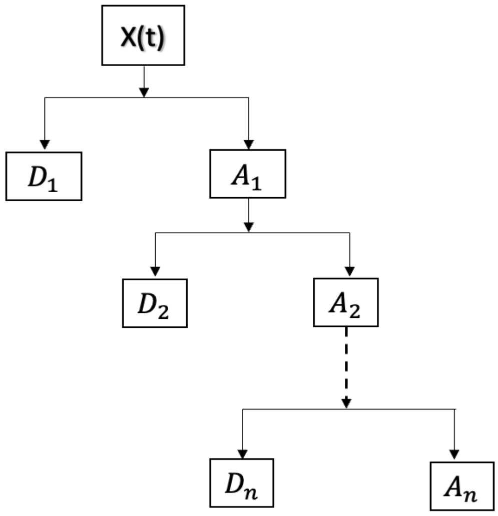

Figure 2.4 : n-level MRWA scheme.

3. EXPERIMENTAL SETUP

Since the vibration data is the best choice for the bearing aging analysis or the bearing FD applications [69], in this study, an experimental setup based on the accelerated aging process is benefited to achieve the vibration signals of an IM. To collect the experimental vibration data, a 5 HP, three-phase and four poles IM is subjected to Electrical Discharge Machining (EDM), thermal aging, and chemical aging processes to realize artificial aging. More information regarding EDM and the Thermal-Chemical Aging processes has been noted in detail in the Ph.D. thesis of Dr. D. Bayram [72].

After executing the EDM process on the motor, the chemical and thermal aging operations were applied to the motor repeatedly, and it was implemented several times to obtain the aged condition. As a result, seven aging cycles are obtained while the aged condition exhibiting the defined vibration thresholds is met. In addition, an initial cycle collected from the healthy motor is also considered to be the first cycle of the dataset. Thus, the acquired dataset consists of eight vibration signals overall. In this study, the other collected electrical and mechanical signals are not used.

In the executed accelerated aging experiment, the accelerometers are located in six different locations on the motor, and only the vibration data collected from one sensor is used eventually. Furthermore, In this experiment, the nominal speed of the IM is 1742 rotation per minute (RPM) with a 60 Hz supply frequency. The data acquisition system consists of a patch panel and a NI SCXI as the signal conditioning interface. NI SCXI-1142 8-channel elliptic low-pass filter module is also used for the conditioning of the vibration signals. In order to prevent high-frequency noise interference, an anti-aliasing filter with a 4 kHz cut-off frequency is employed. Moreover, the vibration data is collected at the 12 kHz sampling frequency and recorded for 10 seconds.

Some of the basic statistical parameters of the healthy cycle and the aged cycles are demonstrated in Table 3.1. The statistical features of each cycle are expected to change with gradually growing aging. Indeed, as seen in Table 3.1, by increasing the bearing aging level, the maximum value, the standard deviation, and the variance of the recorded signal for each cycle are increased accordingly.

Table 3.1 : Statistical features of all aging cycles.

Aging Cycles	Mean	Max	Standard Deviation	Variance
Cycle #1 (initial)	0.0016	0.6009	0.1135	0.0129
Cycle #2	0.0014	0.7988	0.1553	0.0241
Cycle #3	0.0012	0.9991	0.2082	0.0433
Cycle #4	0.0013	1.1970	0.2894	0.0837
Cycle #5	0.0009	1.3240	0.3275	0.1073
Cycle #6	0.0010	1.5658	0.3548	0.1259
Cycle #7	0.0005	1.8516	0.4147	0.1720
Cycle #8	0.0030	2.6113	0.6040	0.3648

Furthermore, to share the differences between each aged cycle in both time and frequency domains, the time and frequency-domain representations of the eight recorded cycles are illustrated in Figure 3.1. Since the PSD is an appropriate tool for analyzing the vibration signals, the PSDs of all cycles are demonstrated as the frequency-domain representation of the acquired signals. The frequency components for the aged cases are amplified intensely between 2 kHz and 4 kHz, compared to the healthier vibration signals. Supportingly, it is indicated that the bearing aging reveals itself on 2-4 kHz in the literature [73]. Also, as expected, the vibration amplitude increases noticeably with aging regarding the time-domain analysis. These results seem to be consistent and meaningful for both time and frequency domain characteristics.

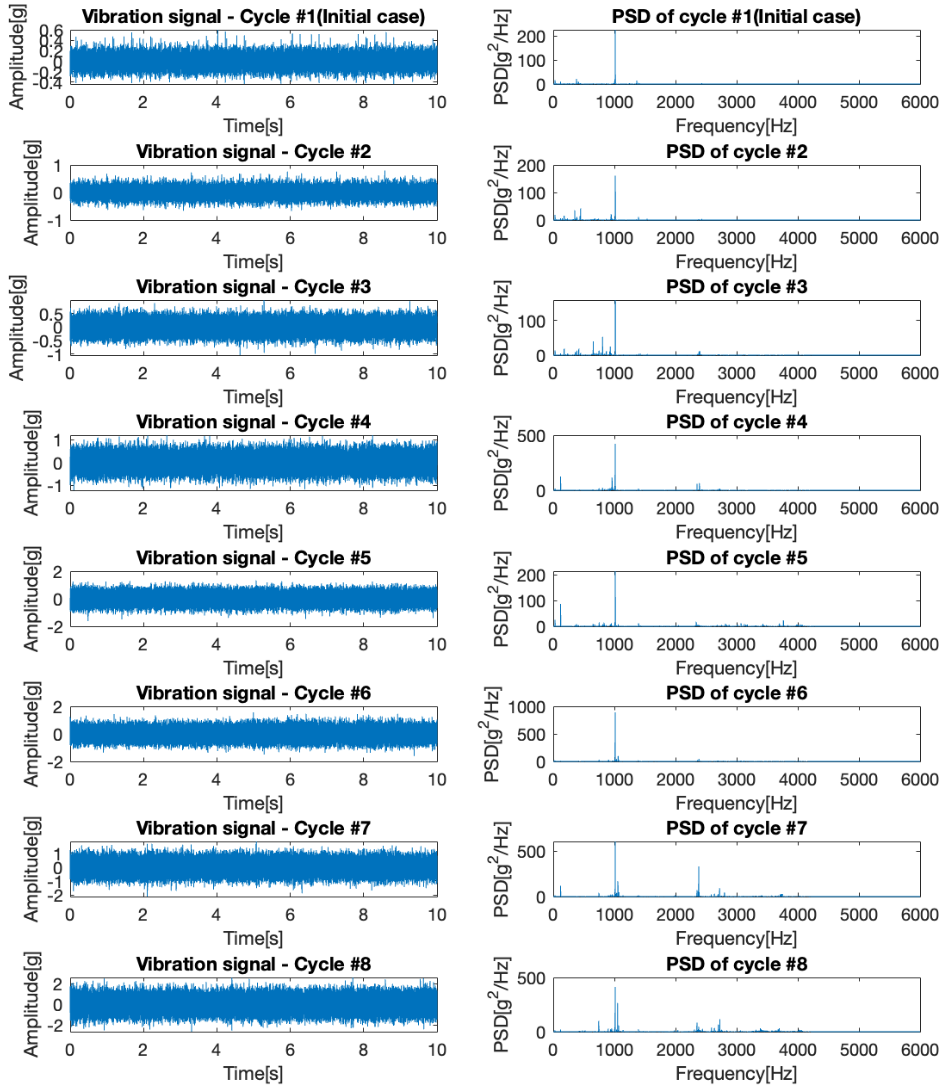

Figure 3.1 : Representation of the recorded vibration signals in both time and frequency domains.

4. NEURAL CLASSIFIERS RATING THE OVERALL CONDITION OF AN ELECTRIC MOTOR

In this section, a PRNN trained with both Levenberg-Marquardt and Scaled Conjugate Gradient algorithms and a Multi-layer FFNN trained with the L-BFGS algorithm are represented to classify the overall condition of an IM, respectively. Both approaches are executed under the same conditions in terms of the type and size of the vibration data sets, the number of splits for each aging cycle, and the feature extraction processes. Besides, in this section, all the NN applications are accomplished by using just 75% of the experimental vibration data.

The major challenge for the training of the proposed ANNs is the selection of suitable features as the input parameters of the network, which may complicate the network structure. It is necessary to select the correct features to increase the performance of the network and reduce the network input dimensions and training time [74].

One of the most popular techniques for extracting the specific features is to calculate the statistical parameters of each data set. There are many research works in which the most efficient time-domain statistical parameters are determined as the inputs of the ANN for CM and FD purposes in IMs [27,28,31].

In this section, nine statistical features are calculated as the input parameters of the proposed ANNs. Indeed, both shallow and deep classifiers are driven by the time-domain statistical features. Mean value, maximum value, kurtosis, skewness, variance, standard deviation, Root Mean Squared (RMS) value, peak-to-peak amplitude, and crest factor are the nine statistical features that have been calculated for each data set. Furthermore, it is crucial to notice that there are eight experimental vibration signals representing the eight conditions of the motor in the data set, and each aging cycle is divided into 40 splits before the feature extraction process.

Table 4.1 indicates the mathematical representation of the employed statistical features. In Table 4.1, x_i is defined as the data samples and N is the total number of samples for each data set.

Table 4.1 : Mathematical representation of the statistical parameters.

Statistical Parameters	Equations
Mean value	$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
Maximum value	$x_{max} = \max\{x_i\}$
Kurtosis	$\overline{x}^4 = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^4}{N} \times \frac{1}{(\delta^2)^2}$
Skewness	$\overline{x}^3 = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^3}{N} \times \frac{1}{(\delta^2)^{\frac{3}{2}}}$
Variance	$\delta^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2$
Standard deviation	$\delta = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}$
Root mean squared (RMS) value	$x_{rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert^2}$
Peak-to-peak amplitude	$x_{peak-to-peak} = \max\{x_i\} - \min\{x_i\}$
Crest factor	$C = \frac{\lvert x_{max}\rvert}{x_{rms}}$

4.1 Pattern Recognition Neural Network (PRNN) to Rate the Condition of an Electric Motor

A PRNN model is developed on MATLAB software to classify the eight different aging levels representing the health conditions of an IM. After extracting the nine statistical features as the inputs of the proposed network, the data in the matrix of the features are normalized between -1 and 1 by using the $Mapminmax$ function, which is provided in the MATLAB Toolbox. Afterward, the matrix of the normalized data is divided into training, validation, and testing subsets with the ratio of 70%, 15%, and 15%, respectively. The proposed PRNN is trained with the BP algorithm and comprises one hidden layer. Levenberg-Marquardt and Scaled Conjugate Gradient optimization algorithms are applied to the proposed network to be assessed separately. Besides, for both types of training methods, two different transfer functions are employed in the hidden layer. The logistic-sigmoid and the tangent-sigmoid are the two activation functions that are selected to quantify the outputs of the hidden layer. The mathematical representations of the logistic-sigmoid and the tangent-sigmoid activation functions are shown in equation (4.1) and equation (4.2), respectively. This application is implemented directly on the acquired experimental vibration data without performing any preprocessing operation. Ultimately, the designed network can be tested with a new data set after executing the training process, and the eight aging levels of the IM can also be classified with relatively high accuracy rates.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{4.1}$$

$$f(x) = \frac{2}{1 + e^{-2x}} - 1 \tag{4.2}$$

In this section, the aim is mainly to evaluate the reliability and performance of the proposed classifiers in rating the overall aging conditions of an IM. For this aim, accuracy rate, Recall, Precision, and F1-score are employed as the significant metrics for the performance evaluation in the proposed multi-class classification applications. Accuracy rate can be defined as the proportion of the total number of correctly estimated samples to the total number of samples availed in the classification. Before introducing the other performance evaluation metrics, four metrics used in a confusion matrix are defined primarily as below:

- True Positive (TP) refers to the number of predictions where the classifier correctly predicts the positive class as positive.
- True Negative (TN) refers to the number of predictions where the classifier correctly predicts the negative class as negative.
- False Positive (FP) refers to the number of predictions where the classifier incorrectly predicts the negative class as positive.
- False Negative (FN) refers to the number of predictions where the classifier incorrectly predicts the positive class as negative.

The Sensitivity or Recall is known as the True Positive Rate (TPR) and determines what fraction of all actual positive samples are correctly predicted as positive by the classifier. Precision assigns what fraction of predictions as a positive class are actually positive. Besides, F1-score is defined as the harmonic mean of Recall and Precision. The mathematical representations of Recall, Precision, and F1-score are represented in equation (4.3), equation (4.4), and equation (4.5), respectively.

$$Recall = \frac{TP}{TP + FN} \tag{4.3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4.4}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4.5}$$

In this study, Recall, Precision, and F1-score values are calculated for each class primarily. Afterward, a total value for each metric in each classifier is also computed. The total Recall value is computed as the sum of TP values across all classes divided by the sum of TP values and FN values across all classes. The total Precision in multi-class classification application is computed by the sum of TP values across all classes divided by the sum of TP values and FP values across all classes.

Besides, the training, validation, testing, and total accuracy rates of the proposed PRNN trained with the Levenberg-Marquardt and the Scaled Conjugate Gradient algorithms are also shown in Table 4.2. It must be noted that the Levenberg-Marquardt and the Scaled Conjugate Gradient algorithms are represented as $trainlm$ and $trainscg$ functions on MATLAB, respectively. By comparing the estimated output results determined in Table 4.2, the PRNN driven by the nine hidden neurons provided the maximum total accuracy rate, which is 99.2%. Indeed, this network is trained with the Levenberg-Marquardt BP algorithm and contains one tangent-sigmoid transfer function in the hidden layer. The testing accuracy rate of the proposed network, which resulted in the maximum total accuracy rate, is also 97.2%. The accuracy rates and the structure of the proposed neural classifier are written bold in Table 4.2.

On the other hand, it is seen that two structures of the developed PRNN produce the maximum testing accuracy rate, which is 100%. The first network is designed with the Levenberg-Marquardt algorithm, nine hidden neurons, and the logistic-sigmoid activation function, and the second network is trained with the Scaled Conjugate Gradient algorithm, nine hidden neurons, and the logistic-sigmoid activation function. The testing accuracy rates of these networks seem to be better than the network in which the testing accuracy rate is 97.2%. However, this value as the testing accuracy rate might not be credible enough in the PRNN applications since it is an ideal value as the testing accuracy rate. Moreover, there are some limitations in terms of the size and type of the experimental vibration data availed in this study. Indeed, the recorded experimental vibration data comprises a pretty small size. Therefore, it is expected to result in an ideal testing accuracy rate by some structures. In addition, it seems that the level of noise that exists in the provided vibration data is very low. It means that the data sets might contain very clear signals, which makes the training process very easy for the proposed network. Thereby, it is more reliable to not choose the network structures with the ideal testing accuracy rates as the optimum classifiers in the PRNN applications.

In addition, there is another network designed with the Levenberg-Marquardt algorithm, four hidden neurons, and the logistic-sigmoid activation function, for which the testing and total accuracy rates are calculated as 94.44% and 98.3%, respectively. The proposed network seems to be also a reliable network [75]. However, the training accuracy rate of the proposed network is also 100%, which might represent the possibility of overfitting in this structure. Furthermore, the network trained with nine hidden neurons and tangent-sigmoid activation function outperforms the structure conducted with four hidden neurons and logistic-sigmoid activation function.

Table 4.2 : Table of the accuracy rates for the PRNN trained with the Levenberg-Marquardt and the Scaled Conjugate Gradient algorithms.

Network Number	Number of Hidden Neurons	Training Function	Transfer Function of Hidden Layer	Accuracy Rates (%)			
				Training	Validation	Testing	Total
#1	10	trainlm	log-sig	98.2	94.4	97.2	97.5
#2	9	trainlm	log-sig	98.8	94.4	100	98.3
#3	8	trainlm	log-sig	99.4	100	86.1	97.5
#4	7	trainlm	log-sig	97.6	80.6	94.4	94.6
#5	6	trainlm	log-sig	99.4	91.7	97.2	97.9
#6	5	trainlm	log-sig	100	94.4	94.4	98.3
#7	4	trainlm	log-sig	100	94.4	94.4	98.3

Table 4.2 (continued) : Table of the accuracy rates for the PRNN trained with the Levenberg-Marquardt and the Scaled Conjugate Gradient algorithms.

Network Number	Number of Hidden Neurons	Training Function	Transfer Function of Hidden Layer	Accuracy Rates (%)			
				Training	Validation	Testing	Total
#8	3	trainlm	log-sig	93.5	86.1	91.7	92.1
#9	10	trainlm	tan-sig	97.6	97.2	91.7	96.7
#10	**9**	**trainlm**	**tan-sig**	**100**	**97.2**	**97.2**	**99.2**
#11	8	trainlm	tan-sig	100	100	88.9	98.3
#12	7	trainlm	tan-sig	100	83.3	91.7	96.3
#13	6	trainlm	tan-sig	97.6	97.2	94.4	97.1
#14	5	trainlm	tan-sig	100	91.7	97.2	98.3
#15	4	trainlm	tan-sig	97.6	94.4	91.7	96.3
#16	3	trainlm	tan-sig	98.8	88.9	97.2	97.1
#17	10	trainscg	log-sig	94	91.7	94.4	93.8
#18	9	trainscg	log-sig	98.8	94.4	100	98.3
#19	8	trainscg	log-sig	98.8	100	91.7	97.9
#20	7	trainscg	log-sig	97.6	91.7	94.4	96.3
#21	6	trainscg	log-sig	96.4	94.4	97.2	96.3
#22	5	trainscg	log-sig	84.5	72.2	75	81.3
#23	4	trainscg	log-sig	96.4	91.7	91.7	95
#24	3	trainscg	log-sig	69	72.2	80.16	71.3
#25	10	trainscg	tan-sig	97.6	94.4	94.4	96.7
#26	9	trainscg	tan-sig	97.6	97.2	97.2	97.5
#27	8	trainscg	tan-sig	97	100	88.9	96.3
#28	7	trainscg	tan-sig	97	100	97.2	97.5
#29	6	trainscg	tan-sig	97	94.4	94.4	96.3
#30	5	trainscg	tan-sig	82.7	75	77.8	80.8
#31	4	trainscg	tan-sig	96.4	94.4	91.7	95.4
#32	3	trainscg	tan-sig	96.4	91.7	97.2	95.8

In order to compare the reliability of the proposed networks containing the high accuracy rates, the performance plots of the four aforementioned networks are illustrated in Figure 4.1. As seen in Figure 4.1, it can be perceived that the PRNN, which is designed with nine hidden neurons, a Levenberg-Marquardt algorithm, and a tangent-sigmoid activation function, has eventuated to the best Mean Squared Error (MSE) value for the validation and training processes. Moreover, it is also demonstrated that in this network, the validation loss decreases until the end of the training process, and the network is not stopped due to the increment of the validation loss. However, in the other structures, the training process ends early when the validation loss reaches the minimum value or exceeds the minimum value. Owing to these facts, compared to the other network designs, the PRNN trained with the nine hidden neurons (given in Figure 4.1 (a)), Levenberg-Marquardt algorithm, and tangent-sigmoid activation function can be verified as the most reliable network structure for the age classification of an IM.

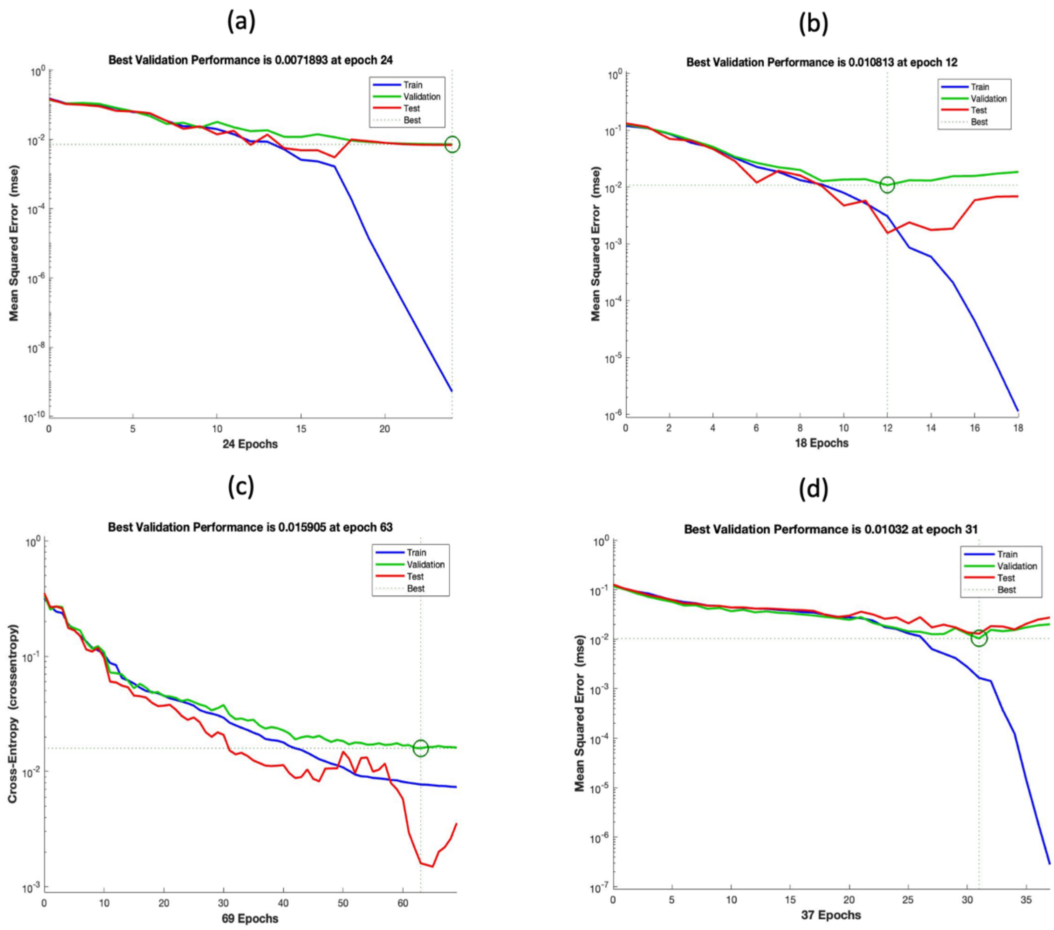

Figure 4.1 : Performance representations of the four different structures of the PRNN defined as network: (a) #10, (b) #2, (c) #18, and (d) #7.

In this section, the output performance of the optimum classifier is also analyzed by the confusion matrix, Receiver Operating Characteristic (ROC), and the error histogram plots. A confusion matrix is a significant performance measurement for classification applications. In this study, the confusion matrix of the optimum classifier for training, validation, testing, and all samples of the experimental vibration data is calculated separately to ensure the consistency of the accuracy rates. The confusion matrix plots of the optimum classifier are illustrated in Figure 4.2. According to these plots, the confusion matrices indicate the approval of the accuracy rates of the optimum network for the aging classification problem. Indeed, it is ascertained that there is one misclassified sample in the validation and testing confusion matrices. Therefore, the confusion matrix for all samples of the experimental data determines two incorrect estimated samples, which certifies the value of the total accuracy rate as 99.2%.

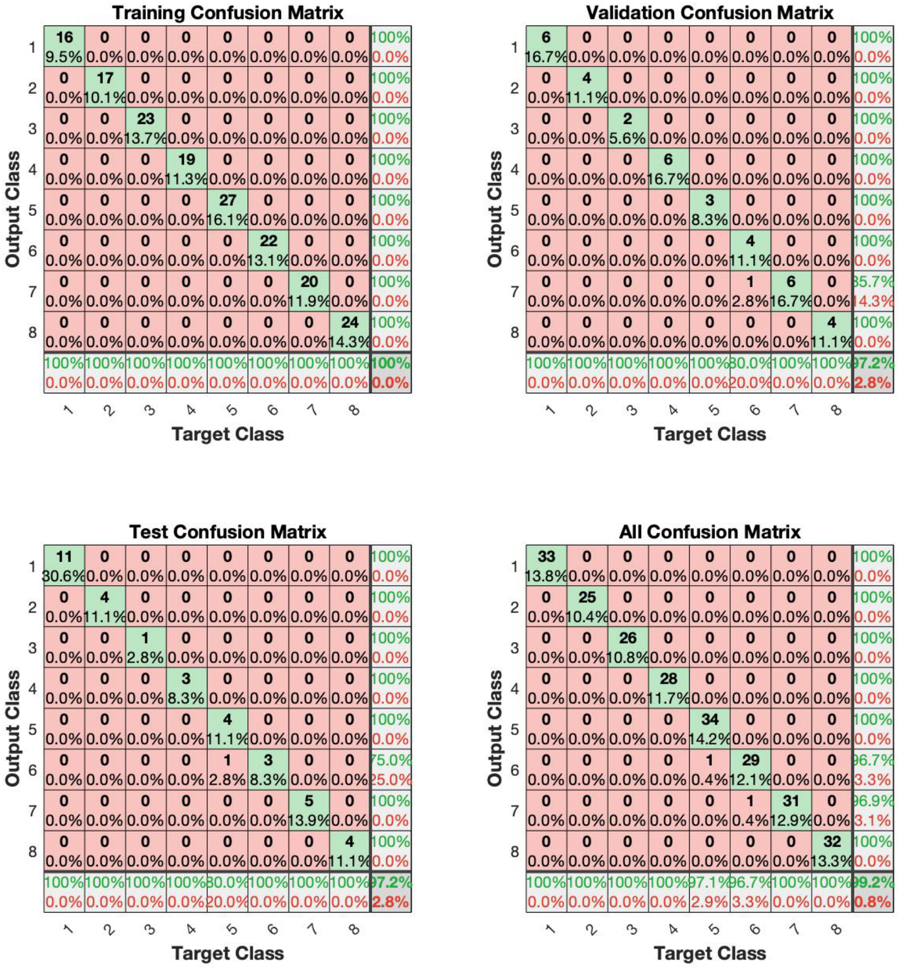

Figure 4.2 : Confusion matrix of the training, validation, testing, and all samples of the vibration data for the network #10 in the PRNN application.

Recall, Precision, and F1-score metrics for each class of the testing process of network #10 are calculated in Table 4.3. as seen in Table 4.3, the Recall value for class 5 and the Precision value for class 6 are 0.8 and 0.75, respectively. Recall, Precision, and F1-score metrics for other classes are calculated as one, which represents a classifier with a perfect performance in classifying the associated classes. These calculations approve that there is just one misclassified sample in the testing confusion matrix, which actually belongs to class 5 but is predicted as class 6. Besides, the calculated F1-score metric for class 5 and class 6 are 0.88 and 0.85, respectively. Indeed, the total values of Recall, Precision, and F1-score are also calculated as 0.972, which certifies the correctness of the testing accuracy rate of network #10.

Table 4.3 : Recall, Precision, and F1-score metrics for the testing process of the network #10 in the PRNN application.

Number of classes	Recall	Precision	F1-score
Class 1	1	1	1
Class 2	1	1	1
Class 3	1	1	1
Class 4	1	1	1
Class 5	0.8	1	0.88
Class 6	1	0.75	0.85
Class 7	1	1	1
Class 8	1	1	1
Total	0.972	0.972	0.972

The ROC curves are the other performance assessment tools for classification purposes. The ROC curves calculate the TPR and the False Positive Rate (FPR) with different values of thresholds for the model. A perfect classifier is expected to have a TPR = 1 and an FPR = 0. Figure 4.3 represents the ROC curves of the training, validation, testing, and all samples of the experimental vibration data for the classifier designed with the Levenberg-Marquardt algorithm, nine hidden neurons, and tangent-sigmoid activation function. As seen in Figure 4.3, the training ROC curves for all classes show the perfect values of TPR and FPR. However, the ROC curves of class 6 and class 7 and the ROC curves of class 5 and class 6 are pulled down in the validation and testing processes, respectively. Nevertheless, for the classification of all samples of the experimental data, acceptable values of TPR and FPR are acquired, and the ROC curves approach the left and top edges of the plot, which proves that the classifier is still operating very well.

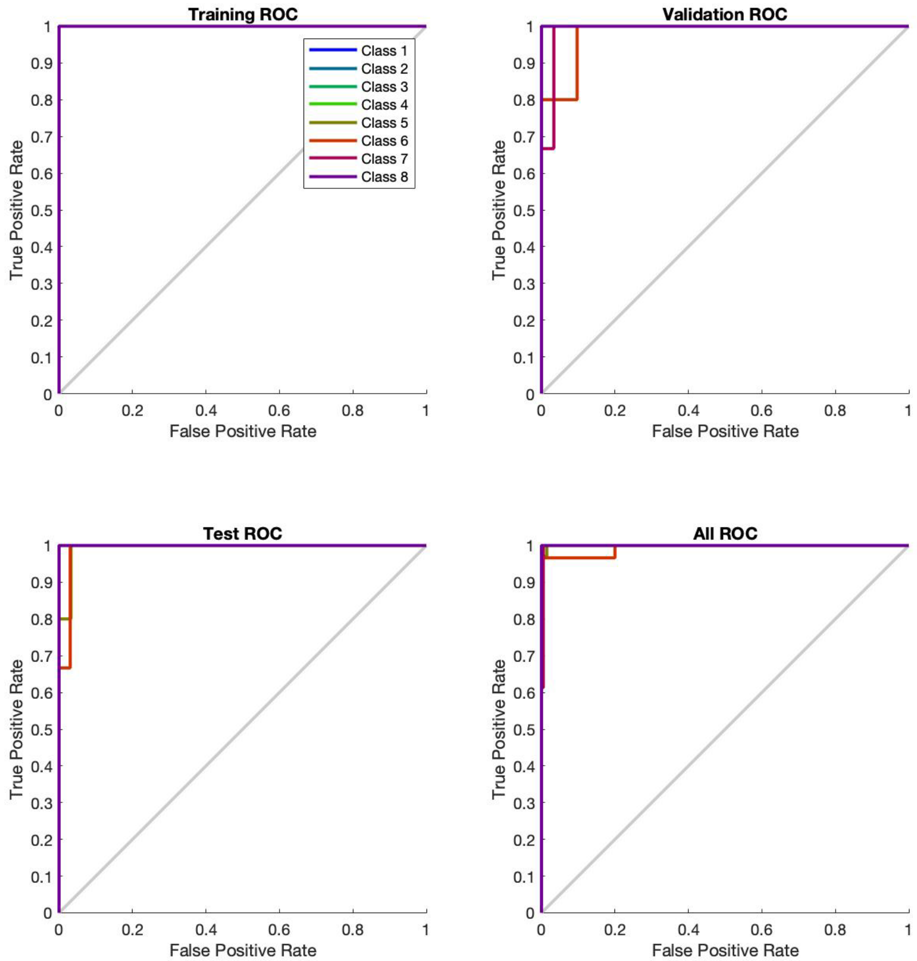

Figure 4.3 : ROC plots of the training, validation, testing, and all samples of the vibration data for the network #10 in the PRNN application.

The error histograms are also suitable indicators for performance evaluation in classification applications. Indeed, the error values of the training, validation and testing processes are determined clearly by using the error histogram plots. The error histograms of the PRNN designed with the Levenberg-Marquardt algorithm, nine hidden neurons, and tangent-sigmoid activation function are also given in Figure 4.4. According to the error histograms, the error value that corresponds to most samples of the data is -0.05. Besides, the total calculated error value of the optimum network is a very small value indicating the high performance of the developed PRNN.

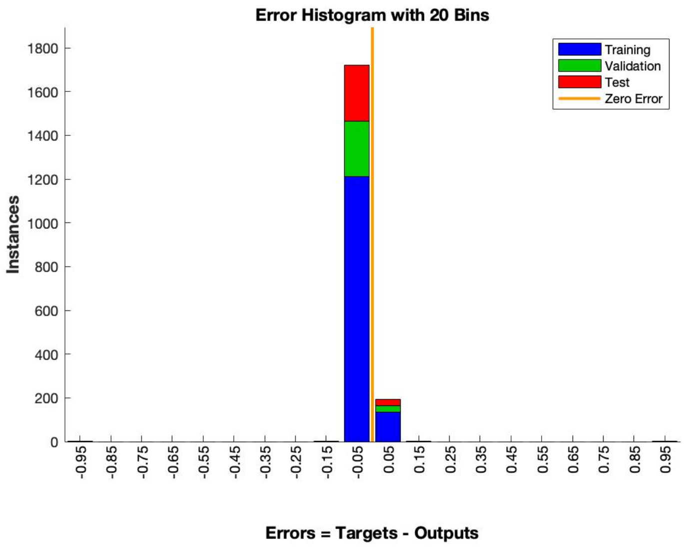

Figure 4.4 : Training, validation, and testing error histogram plots for the network #10 in the PRNN application.

4.2 Multi-Layered FFNN Trained by L-BFGS to Rate the Condition of an Electric Motor

In this section, a fully connected multi-layered FFNN solved by the L-BFGS algorithm is designed as a deep classifier to rate the condition of EM. Indeed, the L-BFGS algorithm is an optimization technique to minimize the cross-entropy loss in the proposed FFNN. This section is focused on the reliability and performance of the L-BFGS-based FFNN in the classification application of the eight aging levels of an IM. Therefore, after creating the matrix of the extracted statistical features and using the $Mapminmax$ function to normalize the data between -1 and 1, the data set is divided into two subsets for just training and testing operations. Generally, the data sets are not divided into three subsets in the L-BFGS-based FFNN applications, and the validation set is not defined beside the testing set simultaneously. However, it is attempted to keep the sizes for the training and testing data sets the same in both L-BFGS-based FFNN and PRNN. The developed deep classifier consists of two hidden layers, and in each hidden layer, one activation function is employed. The activation functions availed in this application are the Rectified Linear Unit ($ReLU$) function and the $Sigmoid$ function. The mathematical definition of the $ReLU$ and $Sigmoid$ activation functions are provided in equation (4.6) and equation (4.7), respectively.

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (4.6)$$

$$f(x) = \frac{1}{1 + e^{-x}} \quad (4.7)$$

In this section, similar to the PRNN application, the developed deep neural classifier employs the experimental vibration data, which is not preprocessed before the training starts. 70 % of the data is then used to train the designed FFNN solved by the L-BFGS algorithm. Afterward, the proposed network is tested by the 15 % of the data provided in the matrix of the features. In this regard, the comparison between two different types of ANN can be achieved accurately to rate the overall aging conditions of the IM.

Several structures of the L-BFGS-based FFNN with their testing accuracy rates are represented in Table 4.4. As seen in Table 4.4, the maximum testing accuracy rate is 100% which indicates the high performance and robustness of the developed deep classifier solved by the L-BFGS algorithm. Indeed, in deep FFNNs, when the testing accuracy rate reaches 100%, it can be accepted as a relatively reliable testing accuracy rate. Since, in the developed L-BFGS-based FFNN application, the data sets are partitioned randomly to create a testing and a training data set for validating a statistical model using cross-validation. In this regard, the possibility of overfitting can be decreased. Indeed, it can be noted that the designed network is constructed as a partitioned classification model using the cross-validation technique. The cross-validation is actually presented to prevent overfitting problems in the ANNs. In the NN applications with training and testing data sets, cross-validation can be used to ensure the randomness of the partitioning process. Therefore, the deep classifier in which the testing accuracy rate is 100% can be preferable compared to the PRNN with the same condition. The acceptable network with the maximum testing accuracy rate is also written bold in Table 4.4.

Table 4.4 : Table of the testing accuracy rates for the multi-layered FFNN solved by the L-BFGS algorithm.

Network Number	Number of Hidden Neurons	Training Function	Transfer Function of each Hidden Layer	Testing Accuracy Rate (%)
	Layer 1 - Layer 2		Layer 1 - Layer 2	
#1	4 - 4	L-BFGS	Sigmoid - Sigmoid	97.22
#2	4 - 7	L-BFGS	Sigmoid - Sigmoid	94.44
#3	4 - 10	L-BFGS	Sigmoid - Sigmoid	100

Table 4.4 (continued) : Table of the testing accuracy rates for the multi-layered FFNN solved by the L-BFGS algorithm.

Network Number	Number of Hidden Neurons	Training Function	Transfer Function of each Hidden Layer	Testing Accuracy Rate (%)
	Layer 1 - Layer 2		Layer 1 - Layer 2	
#4	5 - 3	L-BFGS	Sigmoid - Sigmoid	88.88
#5	5 - 9	L-BFGS	Sigmoid - Sigmoid	83.33
#6	**6 - 6**	**L-BFGS**	**Sigmoid - Sigmoid**	**100**
#7	8 - 4	L-BFGS	Sigmoid - Sigmoid	72.22
#8	7 - 5	L-BFGS	Sigmoid - Sigmoid	94.44
#9	7 - 9	L-BFGS	Sigmoid - Sigmoid	97.22
#10	8 - 6	L-BFGS	Sigmoid - Sigmoid	86.11
#11	9 - 4	L-BFGS	Sigmoid - Sigmoid	91.66
#12	6 - 5	L-BFGS	ReLU - ReLU	83.33
#13	7 - 5	L-BFGS	ReLU - ReLU	97.22
#14	7 - 7	L-BFGS	ReLU - ReLU	97.22
#15	7 - 8	L-BFGS	ReLU - ReLU	86.11
#16	8 - 7	L-BFGS	ReLU - ReLU	88.88
#17	9 - 7	L-BFGS	ReLU - ReLU	94.44
#18	10 - 6	L-BFGS	ReLU - ReLU	94.44
#19	10 - 9	L-BFGS	ReLU - ReLU	91.66

The testing confusion matrices and the testing ROC plots for four optional structures of the proposed deep classifier are illustrated in Figure 4.5 and Figure 4.6, respectively. The four optional networks are determined as:

- Network #3: the network trained with four neurons in the first hidden layer, ten neurons in the second hidden layer, and a Sigmoid activation function in each hidden layer.
- Network #6: the network trained with six neurons in the first hidden layer, six neurons in the second hidden layer, and a Sigmoid activation function in each hidden layer.
- Network #14: the network trained with seven neurons in the first hidden layer, seven neurons in the second hidden layer, and a ReLU activation function in each hidden layer.
- Network #9: the network trained with seven neurons in the first hidden layer, nine neurons in the second hidden layer, and a Sigmoid activation function in each hidden layer.

According to Figure 4.5 and Figure 4.6, it is indicated that the testing confusion matrices and ROC plots of Network #3 and Network #6 approve the correctness of the calculated maximum testing accuracy rate, which is 100% in both networks. Figure 4.5 also demonstrates the testing confusion matrices of Network #14 and Network #9, in which the one misclassified sample verifies the calculated testing accuracy rate (97.2%).

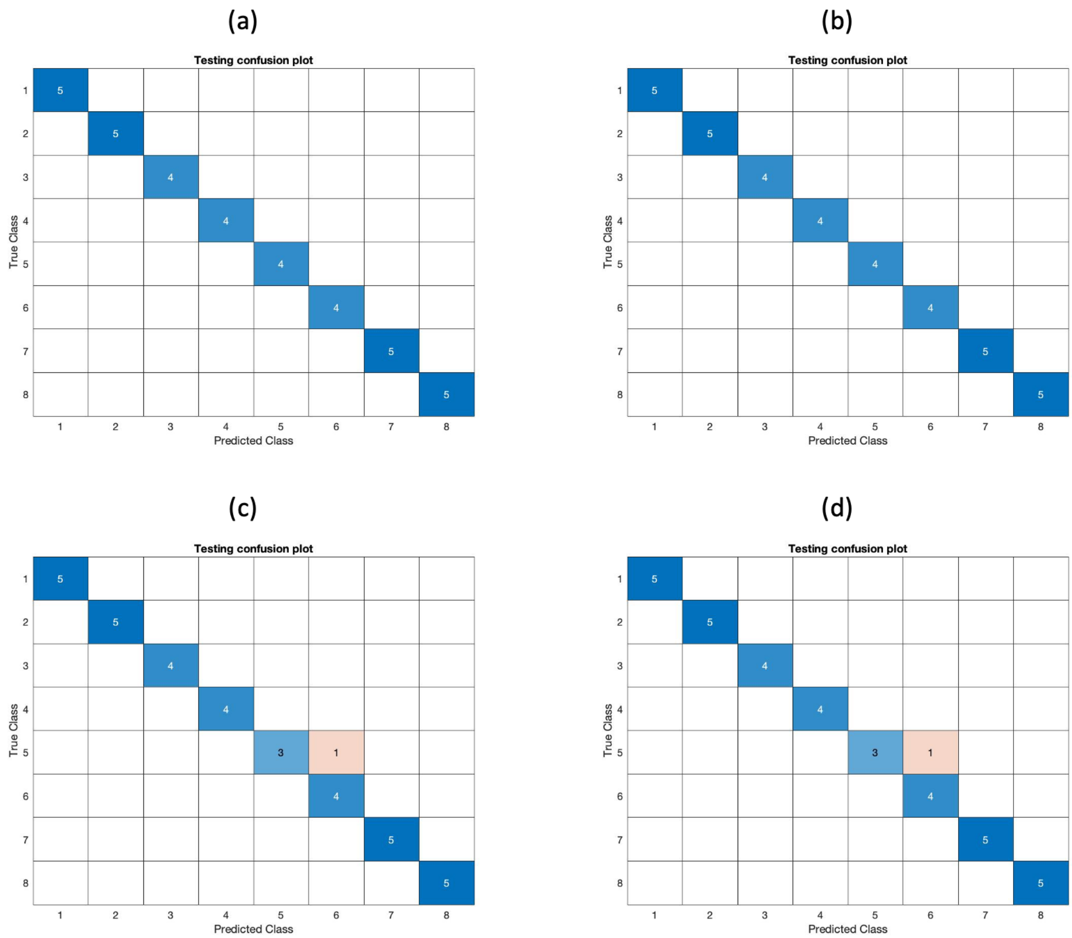

Figure 4.5 : Testing confusion matrices of the developed deep classifier defined as Network: (a) #3, (b)#6, (c) #14, and (d) #9.

The testing ROC plots of Network #14 and Network #9 are also illustrated in Figure 4.6. It can be seen that the ROC curves of class 5 and class 6 are quite distanced from the left and top edges of the plot in Network #14 and Network #9. However, in Network #9, just the ROC curve of class 5 is dragged to the right side of the plot, and the area under the curve for this class is noticeably decreased.

As seen in Table 4.5, Recall, Precision, and F1-score metrics are also calculated for Network #6, which represents the optimum classifier in the L-BFGS-based FFNN application. It is shown that all the metrics for all classes are calculated as one since there is not any misclassified sample in the testing confusion matrix of Network #6. Thereby, the fact of the perfect testing accuracy rate can be verified successfully.

Table 4.5 : Recall, Precision, and F1-score values for the testing process of the Network #6 in the FFNN solved by the L-BFGS algorithm.

Number of classes	Recall	Precision	F1-score
Class 1	1	1	1
Class 2	1	1	1
Class 3	1	1	1
Class 4	1	1	1
Class 5	1	1	1
Class 6	1	1	1
Class 7	1	1	1
Class 8	1	1	1
Total	1	1	1

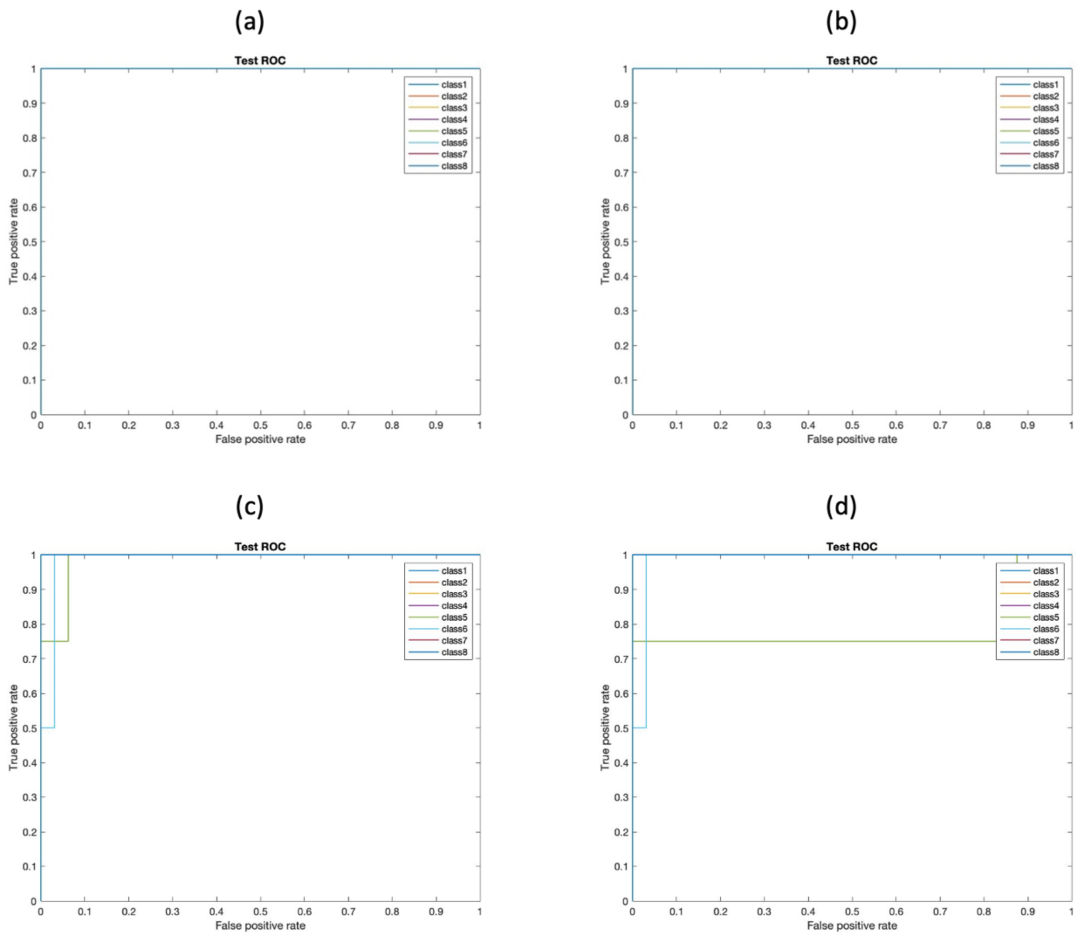

Figure 4.6 : Testing ROC plots of the developed deep classifier defined as Network: (a) #3, (b)#6, (c) #14, and (d) #9.

The training loss plots are also employed as a proper tool for the performance evaluation of the training process in the proposed deep classifier. Figure 4.7 illustrates the training loss plots of the proposed deep classifier for the four optional structures defined earlier. As seen in Figure 4.7, the minimum value of the training loss belongs to Network #6. From this point of view, Network #6 can also be approved as the optimal structure of the developed L-BFGS-based FFNN application.

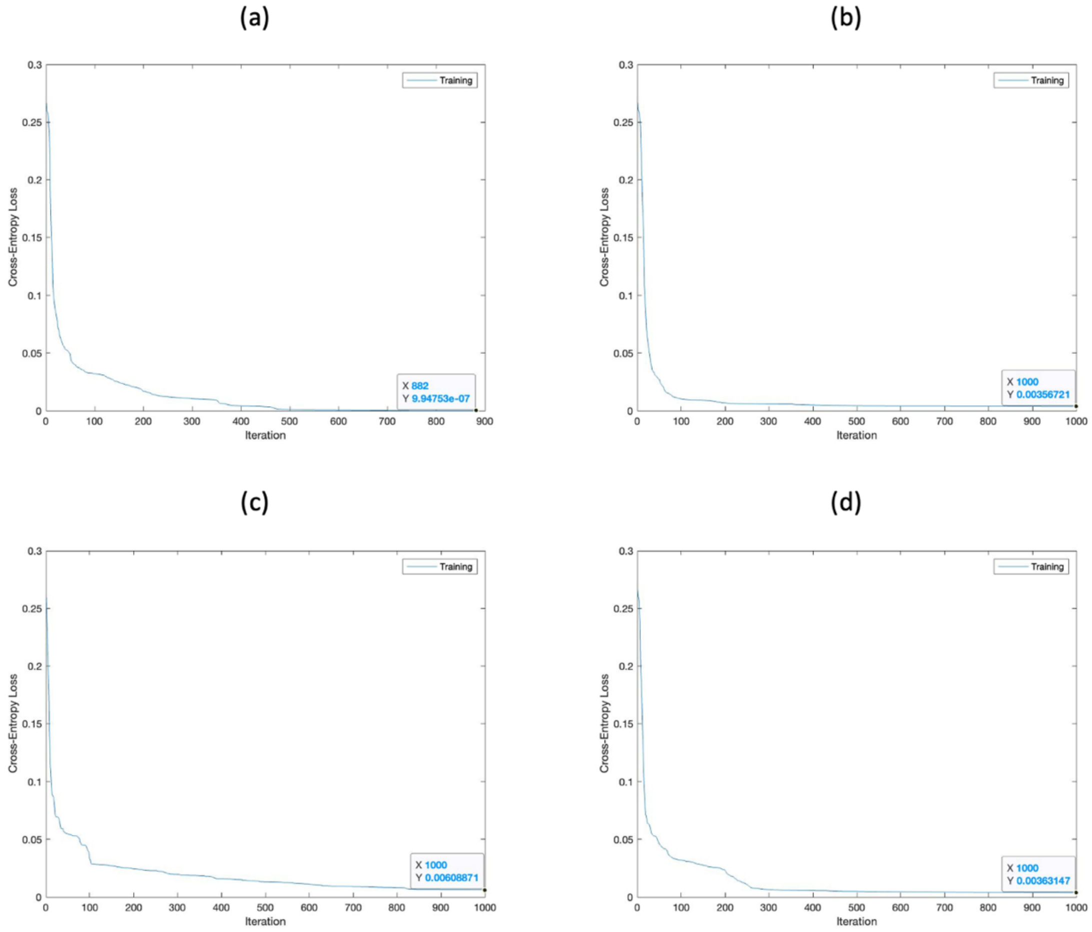

Figure 4.7 : Training cross-entropy loss plots for the four structures of the developed deep classifier defined as Network: (a) #3, (b)#6, (c) #14, and (d) #9.

5. NEURAL CLASSIFIERS RATING THE BEARING DEGRADATION

In this section, two significant changes have been applied to both shallow and deep FFNN applications through the preprocessing and feature selection operations. Indeed, to rate the bearing degradation levels instead of the overall conditions of an IM, the MRWA and a spectral-based feature extraction technique are executed on the whole experimental vibration signals. Therefore, instead of using 75% of the data sets, all the vibration data collected from the experimental setup are employed for the classification of the bearing degradation levels in the IM.

5.1 Preprocessing through MRWA

According to the previous investigations represented in the literature [73], the general characteristic of the bearing degradation can be demonstrated in the high-frequency region of the spectrum, which is determined between 2 and 4 kHz. Thus, the study's primary purpose is to focus on the high-frequency region of the measured vibration signals. Besides, it is preferred to use MRWA as an appropriate tool for decomposing the low and high-frequency bands of the vibration signals. MRWA can extract the desired fault features located in the detail coefficients of each aged cycle accordingly.

In this study, a one-level MRWA is primarily applied to the vibration signals. The Daubechies' 16 (db16), defined in the MATLAB Toolbox, has been employed as the desired mother wavelet function. The approximation and detail coefficients of the proposed MRWA can also denote the low and high-frequency components of the vibration signals as a function of time. Figure 5.1 represents the first detail coefficients of the eight cycles of the measured vibration signals in both time and frequency domains. As seen in Figure 5.1, the characteristics of the first detail coefficients determine the location of the desired bearing fault frequency components (2-4 kHz) in the spectrums. On the other hand, amplitudes of the signals in the time domain are also amplified as the aging level increases. Therefore, the bearing degradation features can be extracted conveniently by using only the first detail of the measured vibration signals.

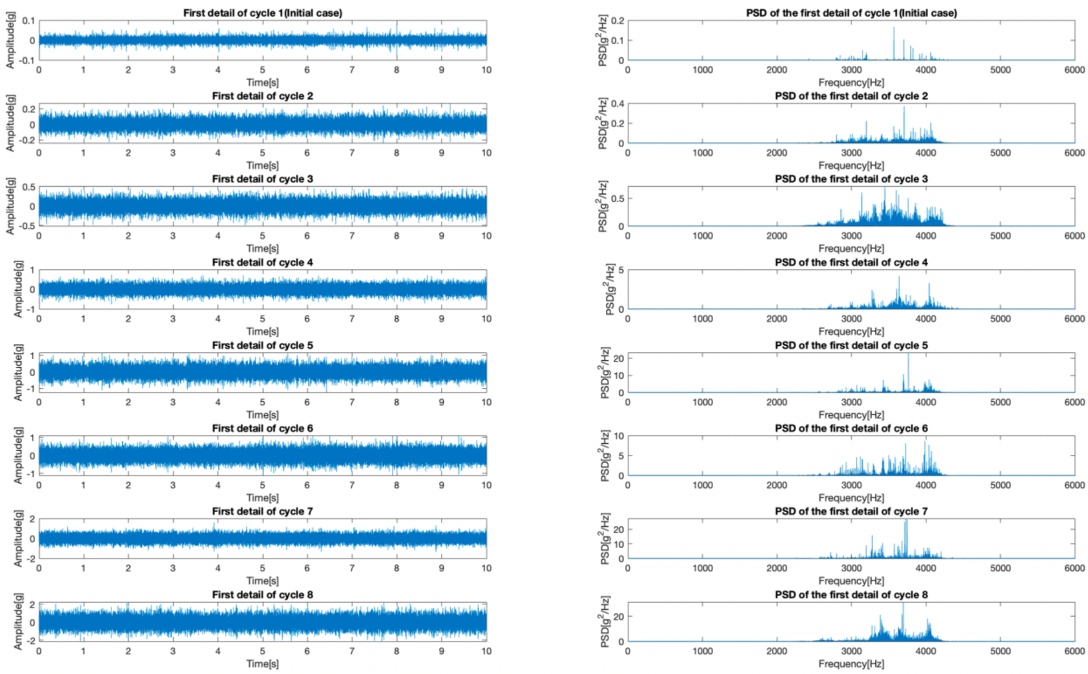

Figure 5.1 : The time-domain and frequency-domain representations of the first detail coefficient of all cycles.

In order to reduce the dimension of the provided data sets, even after MRWA, it is necessary to select some of the discriminant features associated with the bearing degradations in each cycle. In this way, the redundant and unnecessary features will be eliminated efficiently. Indeed, the feature selection process is very advantageous for classification problems in terms of providing high performance and fewer computation efforts in the network. As a result, the learning time will be reduced accordingly by decreasing the complicated calculations in the classifier [76].

There are lots of feature selection methods in the literature. Computing the statistical parameters of the vibration signals is taken into account as a prevalent feature selection method. However, compared to the statistical-based methods, using the spectral-based features is also a superior approach for monitoring the bearing faults located in a specific region. Therefore, some of the significant spectral features which play an essential role in the classification applications are considered to be employed as a novel combination of the spectral-based input parameters. In this study, before computing the five spectral-based features, the eight aging cycles of the vibration data are divided into 40 splits in both shallow and deep MRWA-based applications. The five selected spectral-based features are defined as:

- The Shannon entropy of the signal

- The power of the signal
- The average power of the signal in the 2-4 kHz frequency interval
- The mean frequency
- The percentage of the total power in the 2-4 kHz frequency interval.

The Shannon entropy is created to measure the uncertainty of the random vibration signal. Indeed, the Shannon entropy calculates the expected value of the information which is transferred in a block of data. The units of the computed expected values in Shannon entropy are bits. The mathematical definition of the Shannon Entropy, $H(x)$, for a signal, $x(t)$, is represented in equation (5.1).

$$H(x) = -\sum_{i=1}^{n} P(x_i)\log_2(P(x_i)) \tag{5.1}$$

where i is the number of states ($i = 1,2,\dots,n$) and $P(x_i)$ is the probability of obtaining the value x_i [77].

The power of the vibration signal is calculated by the square of its RMS level. Indeed, there is another definition for the power of a signal in which the squares of the absolute values of the time-domain samples are summed with each other primarily. Afterward, the result of this summation can be divided by the length of the signal. Equation (5.2) represents the mathematical definition associated with the power of a discrete-time signal, $x[n]$.

$$P_x = \frac{1}{N}\sum_{n=0}^{N-1}|x[n]|^2 \tag{5.2}$$

where N is defined as the length of the signal.

The energy of a signal can also be represented by a spectral-based definition. According to Parseval's Theorem, the energy of a signal can be calculated by the sum of the absolute values of the Fourier spectrum coefficients [78]. The mathematical representation of Parseval's Theorem is indicated in equation (5.3).

$$\sum_{n=N}|x[n]|^2 = \frac{1}{N}\sum_{k=N}|x[k]|^2 \tag{5.3}$$

Therefore, to calculate the average power of the signal in the 2-4 kHz frequency interval, the Discrete Fourier Transform (DFT) of the signal, which corresponds to the 2-4 kHz frequency interval, must be calculated primarily. Afterward, the average power can be computed according to the definition of Parseval's Theorem. In this study, the mean frequency is estimated by computing the mean normalized frequency of the power spectrum of the time-domain vibration signals. Eventually, the percentage of the total power in the 2-4 kHz frequency interval, P_{per_power}, is calculated according to the definition represented in equation (5.4).

$$P_{per\_power} = \frac{P_{2-4kHz}}{P_{total}} \times 100 \tag{5.4}$$

where P_{2-4kHz} represents the average power of the signal in the 2-4 kHz frequency interval and P_{total} is defined as the total power of the signal. Both parameters are calculated by integrating the PSD estimate.

5.2 PRNN Trained by the Levenberg-Marquardt Algorithm for Grading Bearing Degradation

A PRNN, which is trained with the Levenberg-Marquardt algorithm, is designed to be evaluated in terms of effectiveness and reliability in bearing aging classification applications. For this aim, a hybrid structure containing both MRWA and PRNN is constructed in this section to grade the bearing degradation in an IM. The PRNN employed in this section includes just one hidden layer. In addition, the performances of the logistic-sigmoid and the tangent-sigmoid activation functions in the hidden layer are also assessed precisely. After obtaining the high-frequency bands (2-4 kHz) and computing the five spectral-based features for each split of the eight aging cycles, the matrix of inputs is prepared to be normalized between 0 and 1. In this step, the normalization technique is implemented by dividing the samples of each column to the maximum value of that column in the matrix of features. This technique is effective for the calculated positive features to be normalized between 0 and 1. Afterward, the data can be divided into three subsets for training, validation, and testing processes with the ratio of 70%, 15%, and 15%, respectively. After the training process, the testing operation is executed to assess the performance of the developed classifier through the accuracy rates which are associated with various network structures. In this application, the Recall, Precision, and F1-score metrics are also calculated to evaluate the performance of the optimum classifier. Besides, the testing accuracy rates of the shallow and deep hybrid classifiers are compared in this section. Table 5.1 determines the training, validation, testing, and total accuracy rates of the proposed hybrid PRNN, which is trained with the Levenberg-Marquardt algorithm. As seen in Table 5.1, the maximum testing accuracy rate is 97.9% which belongs to the network trained with the nine hidden neurons and a logistic-sigmoid activation function. Although there are other structures in which the testing accuracy rate is 97.9%, it is not preferred to select them as the optimal structures since the total accuracy rates associated with these networks are less than 95.6%. Furthermore, compared to the accuracy rates of the PRNN designed in the 4th section, it is evident that the MRWA-based PRNN indicates more reliable performances for the bearing aging classification application. Since the accuracy rates of the proposed hybrid classifier are not 100%, it can be assumed that these bearing aging classification results are reliable enough.

Table 5.1 : Table of the accuracy rates for the MRWA-based PRNN trained with the Levenberg-Marquardt algorithm.

Number of Hidden Neurons	Training Function	Transfer Function of Hidden Layer	Accuracy Rates (%)			
			Training	Validation	Testing	Total
10	trainlm	log-sig	97.3	97.9	89.6	96.2
9	trainlm	log-sig	95.1	95.8	**97.9**	**95.6**
8	trainlm	log-sig	97.3	97.9	89.6	96.2
7	trainlm	log-sig	97.8	97.9	93.8	97.2
6	trainlm	log-sig	96.9	95.8	91.7	95.9
5	trainlm	log-sig	97.3	93.8	89.6	95.6
4	trainlm	log-sig	98.2	97.9	91.7	97.2
3	trainlm	log-sig	96	95.8	89.6	95
2	trainlm	log-sig	87.5	89.6	93.8	88.8
1	trainlm	log-sig	74.6	77.1	72.9	74.7
10	trainlm	tan-sig	99.1	97.9	89.6	97.5
9	trainlm	tan-sig	92.4	97.9	97.9	94.1
8	trainlm	tan-sig	97.3	97.9	89.6	96.2
7	trainlm	tan-sig	97.8	95.8	95.8	97.2
6	trainlm	tan-sig	96.9	95.8	91.7	95.9
5	trainlm	tan-sig	98.7	93.8	87.5	96.2
4	trainlm	tan-sig	97.3	93.8	89.6	95.6
3	trainlm	tan-sig	95.5	95.8	89.6	94.7
2	trainlm	tan-sig	93.3	89.6	97.9	93.4
1	trainlm	tan-sig	75.9	75	70.8	75

The authenticity of the performance for the optimal network, which is trained with the nine hidden neurons and a logistic-sigmoid activation function, is also confirmed by the confusion matrix plots, ROC plots, and the error histogram plots. Figure 5.2, Figure 5.3, and Figure 5.4 demonstrate the confusion matrix plots, ROC plots, and the error histogram plots for the optimal structure of the proposed MRWA-based PRNN, respectively. As seen in Figure 5.2, the training confusion matrix includes 11 misclassified samples associated with the fifth and sixth classes. That is why the training accuracy rate is 95.1%. There are also two misclassified samples for class 5 and class 7 in the validation confusion matrix, which verify the value of the validation accuracy rate. In the testing confusion matrix, just one sample of class 5 is estimated as class 6 by mistake. Therefore, the testing accuracy rate reaches the maximum value by a small number of misclassified samples. Besides, the confusion matrix containing all samples of the vibration data demonstrates the total misclassified samples associated with the training, validation, and testing processes which certifies the total accuracy rate. By analysis of the confusion matrix plots, it can be deduced that the division of the data sets into three subsets has been executed randomly. Generally, the "$Dividerand$" function is availed in the PRNN classification applications. The "$Dividerand$" function is designed in the Toolbox of MATLAB and divides the targets into three sets using the random indices. However, this function is not able to provide a testing or validation data set with a similar number of samples for each class. Indeed, this is an undeniable drawback of this function that impresses the accuracy rates of the PRNNs mostly.

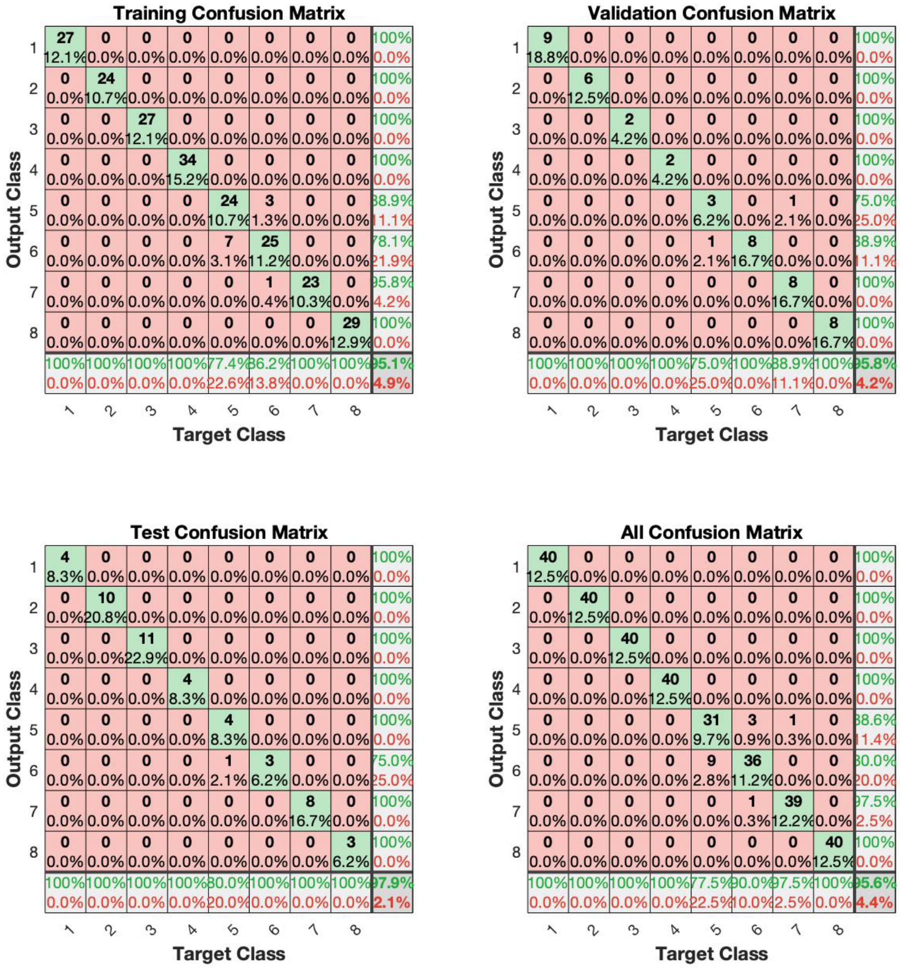

Figure 5.2 : Confusion matrix of the training, validation, testing, and all samples of the vibration data for the optimum structure in the MRWA-based PRNN application.

Recall, Precision, and F1-score metrics for each class of the testing process of the optimum structure are calculated in Table 5.2. As seen in Table 5.2, the Recall value for class 5 and the Precision value for class 6 are 0.8 and 0.75, respectively. Recall, Precision, and F1-score metrics for other classes are calculated as one, which represents a classifier with a perfect performance in classifying the associated classes. These calculations approve that there is just one misclassified sample in the testing confusion matrix, which actually belongs to class 5 but is predicted as class 6. Besides, the calculated F1-score metric for class 5 and class 6 are 0.88 and 0.85, respectively. Indeed, the total values of Recall, Precision, and F1-score are also calculated as 0.979, which represents the correctness of the testing accuracy rate of the optimum classifier.

Table 5.2 : Recall, Precision, and F1-score values for the testing process of the optimum structure in the MRWA-based PRNN.

Number of classes	Recall	Precision	F1-score
Class 1	1	1	1
Class 2	1	1	1
Class 3	1	1	1
Class 4	1	1	1
Class 5	0.8	1	0.88
Class 6	1	0.75	0.85
Class 7	1	1	1
Class 8	1	1	1
Total	0.979	0.979	0.979

By analysis of ROC plots represented in Figure 5.3, it is evident that there are some reductions in the area under the curves associated with some classes. The training ROC plot shows the ROC curves of class 5, class 6, and class 7, which are getting away from the left and top edges of the plot. This reduction just occurred for class 5 and class 6 in the validation process. Moreover, it is also illustrated that the testing ROC curves for all classes create the maximum area under the curve values in the testing process, which indicates the high performance of the proposed hybrid classifier. It can also be perceived that somehow the ROC plot of all samples of the experimental vibration data indicates the combination of the ROC curves in the training, validation, and testing processes for each class.

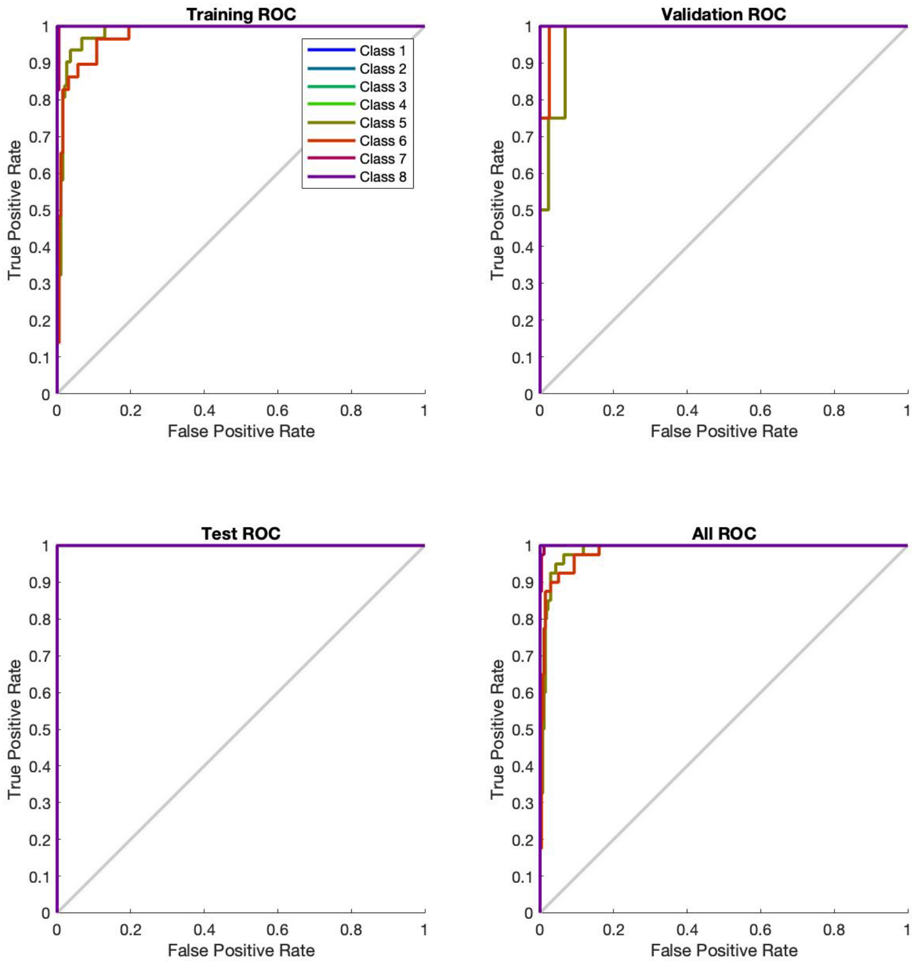

Figure 5.3 : ROC plots of the training, validation, testing, and all samples of the vibration data for the optimal structure in the MRWA-based PRNN application.

As seen in Figure 5.4, the error histogram plot of the proposed PRNN aided by MRWA demonstrates the small error values for training, validation, and testing processes. Indeed, the small error value that corresponds to most data samples gives a significant hint regarding the high performance of the designed classifier. In the training, validation, and testing processes, the error value corresponding to the most data samples is around 0.0165.

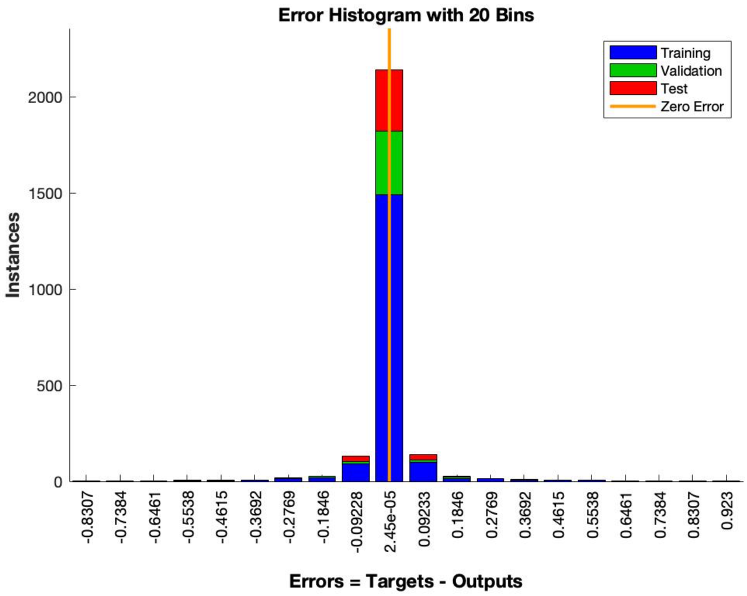

Figure 5.4 : Error histogram plots of the training, validation, and testing processes of the optimal structure in the MRWA-based PRNN application.

5.3 Multi-Layered FFNN Solved by the L-BFGS Algorithm for Grading the Bearing Degradation

In order to classify the bearing aging levels in the IM, a hybrid structure containing a one-level MRWA and a multi-layered FFNN solved by the L-BFGS optimization algorithm is designed in this section. The developed MRWA-based FFNN comprises two hidden layers with two activation functions. The ReLU and the Sigmoid activation functions are considered to be employed in these two hidden layers. Similar to the PRNN application represented in the previous section, the high-frequency bands (2-4 kHz) of the experimental vibration data are primarily decomposed by a one-level MRWA. After splitting the eight cycles into 40 fragments and computing the five spectral-based features for each split, the normalized inputs of the developed hybrid network can be provided. Indeed, the normalization technique is also executed similar to the hybrid PRNN application in which all samples are normalized between 0 and 1. In this multi-layered deep FFNN application, the whole data set is divided again into two subsets by using the partitioning technique, which is used in the cross-validated models. One subset is availed for the training operation with the ratio of 70%, and the other subset is used for the testing process comprising the ratio of 15%. In this application, the size of the training and testing data sets are considered to be similar to the training and testing data sets in the MRWA-based PRNN application, so the remaining 15% of the data is not employed. Eventually, the training and testing processes can be started, respectively. Table 5.3 represents some of the structures that are led to high testing accuracy rates in the L-BFGS-based FFNN application aided by MRWA. According to the testing accuracy rates represented in Table 5.3, it is determined that there are four structures resulting in the maximum testing accuracy rate, which is 100%. The four mentioned networks with the highest testing accuracy rates are determined as:

- Network #5, which represents the network trained with seven neurons in the first hidden layer, six neurons in the second hidden layer, and a sigmoid activation function in each hidden layer.
- Network #8, which represents the network trained with nine neurons in the first hidden layer, seven neurons in the second hidden layer, and a sigmoid activation function in each hidden layer.
- Network #9, which represents the network trained with ten neurons in the first hidden layer, nine neurons in the second hidden layer, and a sigmoid activation function in each hidden layer.
- Network #17, which represents the network trained with nine neurons in the first hidden layer, six neurons in the second hidden layer, and a ReLU activation function in each hidden layer.

Table 5.3 : Table of the testing accuracy rates for the MRWA-based FFNN solved by the L-BFGS algorithm.

Network Number	Number of Hidden Neurons	Training Function	Transfer Function of each Hidden Layer	Testing Accuracy Rate (%)
	Layer 1 - Layer 2		Layer 1 - Layer 2	
#1	3 - 5	L-BFGS	Sigmoid - Sigmoid	97.91
#2	4 - 4	L-BFGS	Sigmoid - Sigmoid	95.83
#3	4 - 6	L-BFGS	Sigmoid - Sigmoid	95.83
#4	4 - 10	L-BFGS	Sigmoid - Sigmoid	93.75
#5	7 - 6	L-BFGS	Sigmoid - Sigmoid	100
#6	7 - 10	L-BFGS	Sigmoid - Sigmoid	97.91
#7	8 - 9	L-BFGS	Sigmoid - Sigmoid	93.75
#8	9 - 7	L-BFGS	Sigmoid - Sigmoid	100
#9	**10 - 9**	**L-BFGS**	**Sigmoid - Sigmoid**	**100**
#10	10 - 10	L-BFGS	Sigmoid - Sigmoid	95.83
#11	4 - 7	L-BFGS	ReLU - ReLU	95.83
#12	4 - 9	L-BFGS	ReLU - ReLU	97.91

Table 5.3 (continued) : Table of the testing accuracy rates for the MRWA-based FFNN solved by the L-BFGS algorithm.

Network Number	Number of Hidden Neurons	Training Function	Transfer Function of each Hidden Layer	Testing Accuracy Rate (%)
	Layer 1 - Layer 2		Layer 1 - Layer 2	
#13	6 - 3	L-BFGS	ReLU - ReLU	97.91
#14	6 - 8	L-BFGS	ReLU - ReLU	95.83
#15	6 - 9	L-BFGS	ReLU - ReLU	97.91
#16	8 - 9	L-BFGS	ReLU - ReLU	95.83
#17	9 - 6	L-BFGS	ReLU - ReLU	100
#18	10 - 10	L-BFGS	ReLU - ReLU	95.83

In order to select one of these four structures as the optimal structure, the training cross-entropy loss plots of each structure are demonstrated in Figure 5.5. By comparing the minimum training loss values associated with each structure, Network #9 can be chosen as the optimal structure. Indeed, the minimum training loss for Network #9 is 0.0142018, which is the significantly small value compared to the other structures. The optimal structure is determined in bold in Table 5.3. Furthermore, it can be determined that the proposed hybrid network outperforms the MRWA-based PRNN application in terms of reliability and performance in the classification of the bearing degradation levels in the IM.

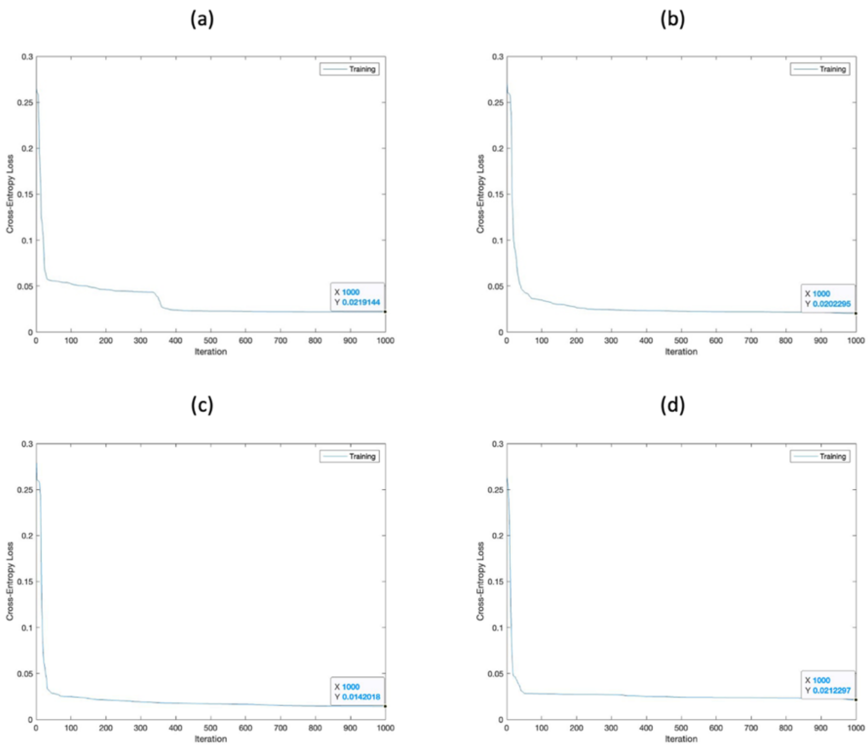

Figure 5.5 : Training cross-entropy loss plots for the four structures of the MRWA-based deep classifier defined as Network: (a) #5, (b) #8, (c) #9, and (d) #17.

Since the testing confusion matrices and the testing ROC plots of the aforementioned four structures are precisely similar to each other, only the testing confusion matrix and the testing ROC plot of the optimum structure are demonstrated in Figure 5.6. Obviously, there is no misclassified sample in the testing confusion matrix. Thereby, the perfect performance of the optimum network can be appropriately justified. Indeed, since the overall size of the collected vibration data set is relatively small in such an application, the testing data set can also be created with a small size. Therefore, by creating some minor variations in the number of misclassified samples, the testing accuracy rates can be changed to a large extent. In other words, when there is just one misclassified sample in the testing confusion matrix, and the associated testing accuracy rate reaches 97.9%, it is natural to reach the perfect testing performance when there is not any misclassified sample in the testing confusion matrix.

Besides, it is determined that the ROC curve of each class hugs the left and top edges of the plot, which represents the high performance of the proposed classifier. As seen in the testing confusion matrix plot, the significant advantage of the multi-layered L-BFGS-based FFNN application is the fact that the testing data set can be provided by a partitioning technique that is implemented randomly. Indeed, an appropriate feature of this technique is the fact that each class includes the same number of samples in the testing data set. Therefore, unlike the PRNN model, which uses the "$Dividerand$" function, in the partitioned classification models, each class of the testing data set includes the same number of samples.

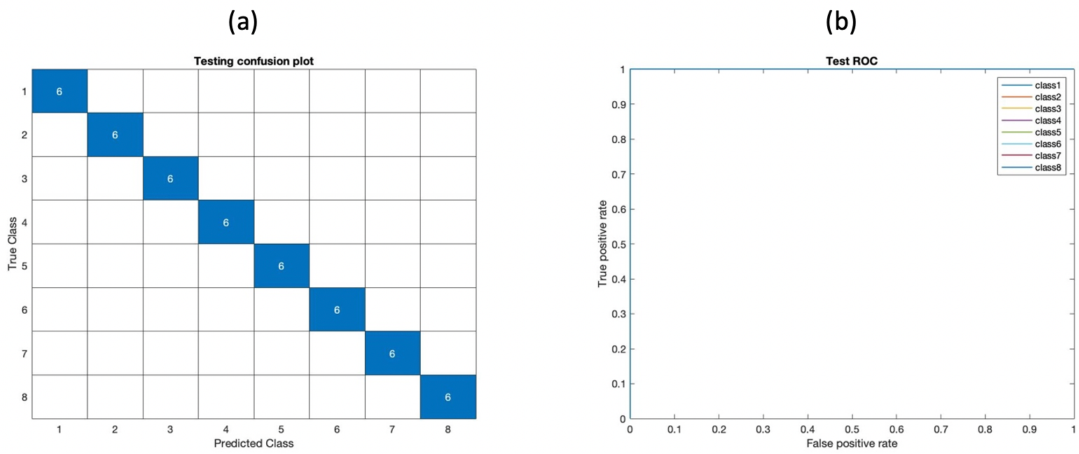

Figure 5.6 : Performance measurement plots for the optimum structure of the MRWA-based FFNN solved by the L-BFGS algorithm: (a) testing confusion matrix and (b) testing ROC plot.

As seen in Table 5.4, Recall, Precision, and F1-score metrics are also calculated for Network #9, which represents the optimum structure of the MRWA-based deep classifier. It is shown that all the metrics for all classes are calculated as one since there is no misclassified sample in the testing confusion matrix of Network #9. Thus, the perfect testing accuracy rate of Network #9 can be verified successfully.

Table 5.4 : Recall, Precision, and F1-score values for the testing process of Network #9 in the MRWA-based FFNN solved by the L-BFGS algorithm.

Number of classes	Recall	Precision	F1-score
Class 1	1	1	1
Class 2	1	1	1
Class 3	1	1	1
Class 4	1	1	1
Class 5	1	1	1
Class 6	1	1	1
Class 7	1	1	1
Class 8	1	1	1
Total	1	1	1

6. CONCLUSIONS

In this thesis, four ANN-based CM approaches are presented to classify the different aging conditions of an IM by using vibration signals, which are the most informative signals for monitoring purposes. In this study, vibration signals are availed not only for rating the overall conditions but also as the best shortcut to classify the bearing aging conditions of IM. In other words, the study consists of two parts; rating the overall condition and the bearing condition of the motor, respectively. Therefore, seven aged vibration signals are collected through an artificially accelerated aging experiment, including EDM and the chemical-thermal aging actions. In addition to the seven aging cycles, an initial cycle is also added to the acquired vibration data sets representing the healthy condition of IM. In this study, four neuro classifiers are represented to grade the different conditions of the IM. The performances of these four neuro classifiers are evaluated through the accuracy rates, Recall, Precision, F1-score, confusion matrices, ROC plots, and loss plots. The error histogram plots are also used to evaluate the PRNN applications. Properties, comments, and findings of each classifier are listed below:

- Primarily, a PRNN model is designed to classify the overall condition of the IM by training the nine time-domain statistical features of the acquired vibration signals. The accuracy rates of the proposed shallow NN trained with different optimization algorithms are compared. These BP-based optimization algorithms are the Levenberg-Marquardt and the Scaled Conjugate Gradient algorithms. It is indicated that the network trained with the Levenberg-Marquardt algorithm outperforms the network trained with the Scaled Conjugate Gradient algorithm. The maximum total accuracy rate of the best structure trained by the Levenberg-Marquardt algorithm is 99.2%, while the testing accuracy rate associated with this network is 97.2%. By analyzing the metrics and comparing the performance measurement plots of different structures, it is derived that even the optimum structure of the PRNN is not reliable enough for the aging classification of the IM. Also, the 100% training accuracy rate, calculated for the optimum structure, represents the possibility of overfitting in the proposed network.

- In order to reduce the risk of memorization and create a model containing more reliable structures, a two-layered FFNN model solved by the L-BFGS algorithm is developed as a deep classifier to rate rating the overall conditions of the IM. This classifier's maximum testing accuracy rate is 100%, which indicates a better performance than the PRNN application. In addition, all the calculated metrics and performance measurement plots certify the precision and correctness of the selected network as an optimum structure. In this application, the maximum testing accuracy rate is still more reliable than the PRNN application since the proposed deep classifier uses a creditable tool for the random division of the testing data set, preventing the overfitting in deep NN applications. However, the ideal testing accuracy rate can also be justified because the size of the experimental vibration data is relatively small, and it is effortless for a deep classifier to learn the patterns efficiently.
- In the third application, a one-level MRWA technique is employed as a preprocessing method to extract the bearing fault regions (2-4 kHz) in the spectrum of the vibration signals since this application is developed to grade the bearing aging levels of the IM. Afterward, five new spectral-based features are proposed as the inputs of a PRNN, which is trained by the Levenberg-Marquardt algorithm. The reliability of the created network is also enhanced by increasing the size of the vibration data sets. By investigating the computed metrics and the performance measurement plots, it is deduced that the total and testing accuracy rates of the optimum structure in the proposed PRNN are 95.6% and 97.9%, respectively. This application is more reliable and realistic since the optimum training accuracy rate is not 100%, and the possibility of memorization is reduced by increasing the size of the data sets.
- Eventually, as the fourth classifier, an MRWA-based FFNN is trained by the L-BFGS algorithm to be evaluated and compared with the proposed MRWA-based PRNN in terms of the reliability and performance in bearing aging classification application. This classification is also based on the five spectral-based features. By analysis of the four computed metrics and the performance measurement plots, it is ascertained that the maximum testing accuracy rate is

100%, which belongs to the network with the minimum training loss. Besides, the quality of the performance for the optimum classifier is investigated appropriately by using the confusion matrix, the ROC, and the training loss plots. By comparing the metrics, the testing confusion matrix, and the testing ROC plots of the optimum structures in both shallow and deep classifiers, it can be figured out that the performance and reliability of the bearing aging classification application are enhanced by using the L-BFGS-based classifier. Therefore, the MRWA-based deep classifier is suggested as the most robust and reliable model for monitoring and detecting the bearing aging level.

It is also essential to state that the proposed study introduces a novel strategy by extracting the spectral-based features and providing an efficient hybrid structure for the aging monitoring applications. By computing the five spectral-based features instead of nine time-domain statistical features as the inputs of the MRWA-based classifiers, the dimensions of the network inputs are also decreased efficiently. Furthermore, the high performance and accuracy scores of the proposed deep neuro classifier can be helpful for industrial applications, which substantially require the CM and FD proceedings to prevent any irreparable damages associated with the bearing degradations in EMs.

REFERENCES

[1] **Hashemnia, N., & Asaei, B.** (2008). Comparative study of using different electric motors in the electric vehicles. *2008 18th International Conference on Electrical Machines*.

[2] **Faiz, J., Ebrahimi, B., & Sharifian, M.** (2006). Different faults and their diagnosis techniques in three-phase squirrel-cage induction motors—A review. *Electromagnetics*, *26*(7): 543-569.

[3] **Seera, M., Lim, C. P., Nahavandi, S., & Loo, C. K.** (2014). Condition monitoring of induction motors: A review and an application of an ensemble of hybrid intelligent models. *Expert Systems with Applications*, *41*(10): 4891-4903.

[4] **Dash, R. N., Sahu, S., Panigrahi, C. K., & Subudhi, B.** (2016). Condition monitoring of induction motors:—a review. *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*.

[5] **Choudhary, A., Goyal, D., Shimi, S. L., & Akula, A.** (2019). Condition monitoring and fault diagnosis of induction motors: A review. *Archives of Computational Methods in Engineering*, *26*(4): 1221-1238.

[6] **Irgat, E., Çinar, E., & Ünsal, A.** (2021). The detection of bearing faults for induction motors by using vibration signals and machine learning. *2021 IEEE 13th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED)*.

[7] **Tavner, P. J.** (2008). Review of condition monitoring of rotating electrical machines. *IET electric power applications*, *2*(4): 215-247.

[8] **Seker, S.** (2000). Determination of air-gap eccentricity in electric motors using coherence analysis. *IEEE Power Engineering Review*, *20*(7): 48-50.

[9] **Werynski, P., Roger, D., Corton, R., & Brudny, J. F.** (2006). Proposition of a new method for in-service monitoring of the aging of stator winding insulation in AC motors. *IEEE Transactions on Energy Conversion*, *21*(3): 673-681.

[10] **Bayram, D., & Şeker, S.** (2015). Anfis model for vibration signals based on aging process in electric motors. *Soft Computing*, *19*(4): 1107-1114.

[11] **Şeker, S., Ayaz, E., Upadhyaya, B., & Erbay, A.** (2000). Analysis of motor current and vibration signals for detecting bearing damage in electric motors. *MARCON 2000 Maintenance and Reliability Conference, Knoxville*.

[12] **Seker, S., & Ayaz, E.** (2003). Feature extraction related to bearing damage in electric motors by wavelet analysis. *Journal of the Franklin Institute*, *340*(2): 125-134.

[13] **Karatoprak, E., Senguler, T., Ayaz, E., Caglar, R., & Seker, S.** (2007). Spectral and statistical based modeling for bearing damage in induction motors. *2007 IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*.

[14] **Gu, F., Wang, T., Alwodai, A., Tian, X., Shao, Y., & Ball, A.** (2015). A new method of accurate broken rotor bar diagnosis based on modulation signal bispectrum analysis of motor current signals. *Mechanical Systems and Signal Processing*, *50*: 400-413.

[15] **Benbouzid, M. E. H., Vieira, M., & Theys, C.** (1999). Induction motors' faults detection and localization using stator current advanced signal processing techniques. *IEEE Transactions on power electronics*, *14*(1): 14-22.

[16] **Zhang, P., Du, Y., Habetler, T. G., & Lu, B.** (2010). A survey of condition monitoring and protection methods for medium-voltage induction motors. *IEEE Transactions on Industry Applications*, *47*(1): 34-46.

[17] **Gritli, Y., Bellini, A., Rossi, C., Casadei, D., Filippetti, F., & Capolino, G.** (2017). Condition monitoring of mechanical faults in induction machines from electrical signatures: Review of different techniques. *2017 IEEE 11th international symposium on diagnostics for electrical machines, power electronics and drives (SDEMPED)*.

[18] **Costa, C. d., Kashiwagi, M., & Mathias, M. H.** (2015). Rotor failure diagnosis of induction motors by wavelet transform and Fourier transform in function of the load. *International Conference on Computer Science and Artificial Intelligence (ICCSAI 2014)*.

[19] **Kia, S. H., Henao, H., & Capolino, G.-A.** (2009). Diagnosis of broken-bar fault in induction machines using discrete wavelet transform without slip estimation. *IEEE Transactions on Industry Applications*, *45*(4): 1395-1404.

[20] **Patel, V. U.** (2019). Condition monitoring of induction motor for broken rotor bar using discrete wavelet transform & K-nearest neighbor. *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*.

[21] **Yavanarani, K., Raj, G. S. S., Christabel, S. S., & Vijayaraghavan, J.** (2010). Multi resolution analysis for bearing fault diagnosis. *Frontiers in Automobile and Mechanical Engineering-2010*.

[22] **Nishat Toma, R., & Kim, J.-M.** (2020). Bearing fault classification of induction motors using discrete wavelet transform and ensemble machine learning algorithms. *Applied Sciences*, *10*(15): 5251.

[23] **Lashkari, N., Poshtan, J., & Azgomi, H. F.** (2015). Simulative and experimental investigation on stator winding turn and unbalanced supply voltage fault diagnosis in induction motors using Artificial Neural Networks. *ISA transactions*, *59*: 334-342.

[24] **Wangngon, B., Sittisrijan, N., & Ruangsinchaiwanich, S.** (2014). Fault detection technique for identifying broken rotor bars by artificial neural

network method. *2014 17th International Conference on Electrical Machines and Systems (ICEMS)*.

[25] **Goundar, S., Pillai, M., Mamun, K., Islam, F., & Deo, R.** (2015). Real time condition monitoring system for industrial motors. *2015 2nd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*.

[26] **Liang, X., & Edomwandekhoe, K.** (2017). Condition monitoring techniques for induction motors. *2017 IEEE Industry Applications Society Annual Meeting*.

[27] **Zarei, J., Tajeddini, M. A., & Karimi, H. R.** (2014). Vibration analysis for bearing fault detection and classification using an intelligent filter. *Mechatronics*, *24*(2): 151-157.

[28] **Zarei, J.** (2012). Induction motors bearing fault detection using pattern recognition techniques. *Expert Systems with Applications*, *39*(1): 68-73.

[29] **Li, B., Chow, M.-Y., Tipsuwan, Y., & Hung, J. C.** (2000). Neural-network-based motor rolling bearing fault diagnosis. *IEEE transactions on industrial electronics*, *47*(5): 1060-1069.

[30] **Al-Raheem, K. F., & Abdul-Karem, W.** (2010). Rolling bearing fault diagnostics using artificial neural networks based on Laplace wavelet analysis. *International Journal of Engineering, Science and Technology*, *2*(6).

[31] **Saravanan, N., Siddabattuni, V. K., & Ramachandran, K.** (2010). Fault diagnosis of spur bevel gear box using artificial neural network (ANN), and proximal support vector machine (PSVM). *Applied Soft Computing*, *10*(1): 344-360.

[32] **Sornmuang, S., & Suwatthikul, J.** (2011). Detection of a motor bearing shield fault using neural networks. *SICE Annual Conference 2011*.

[33] **Farajzadeh-Zanjani, M., Razavi-Far, R., Saif, M., & Rueda, L.** (2016). Efficient feature extraction of vibration signals for diagnosing bearing defects in induction motors. *2016 International Joint Conference on Neural Networks (IJCNN)*.

[34] **Kumar, H., Pai, P. S., Sriram, N., & Vijay, G.** (2013). ANN based evaluation of performance of wavelet transform for condition monitoring of rolling element bearing. *Procedia engineering*, *64*: 805-814.

[35] **Castejón, C., Lara, O., & García-Prada, J.** (2010). Automated diagnosis of rolling bearings using MRA and neural networks. *Mechanical Systems and Signal Processing*, *24*(1): 289-299.

[36] **Janssens, O., Slavkovikj, V., Vervisch, B., Stockman, K., Loccufier, M., Verstockt, S., Van de Walle, R., & Van Hoecke, S.** (2016). Convolutional neural network based fault detection for rotating machinery. *Journal of Sound and Vibration*, *377*: 331-345.

[37] **Moosavian, A., Jafari, S. M., Khazaee, M., & Ahmadi, H.** (2018). A Comparison Between ANN, SVM and Least Squares SVM: Application in Multi-Fault Diagnosis of Rolling Element Bearing. *International Journal of Acoustics & Vibration*, *23*(4).

[38] **Guo, X., Chen, L., & Shen, C.** (2016). Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement*, *93*: 490-502.

[39] **Liu, R., Meng, G., Yang, B., Sun, C., & Chen, X.** (2016). Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine. *IEEE Transactions on Industrial Informatics*, *13*(3): 1310-1320.

[40] **Lu, C., Wang, Z., & Zhou, B.** (2017). Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. *Advanced Engineering Informatics*, *32*: 139-151.

[41] **Xia, M., Li, T., Xu, L., Liu, L., & De Silva, C. W.** (2017). Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks. *IEEE/ASME transactions on mechatronics*, *23*(1): 101-110.

[42] **Wen, L., Li, X., Gao, L., & Zhang, Y.** (2017). A new convolutional neural network-based data-driven fault diagnosis method. *IEEE transactions on industrial electronics*, *65*(7): 5990-5998.

[43] **Zhang, W., Zhang, F., Chen, W., Jiang, Y., & Song, D.** (2018). Fault state recognition of rolling bearing based fully convolutional network. *Computing in Science & Engineering*, *21*(5): 55-63.

[44] **Zilong, Z., & Wei, Q.** (2018). Intelligent fault diagnosis of rolling bearing using one-dimensional multi-scale deep convolutional neural network based health state classification. *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*.

[45] **Zhang, W., Li, C., Peng, G., Chen, Y., & Zhang, Z.** (2018). A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mechanical Systems and Signal Processing*, *100*: 439-453.

[46] **Li, S., Liu, G., Tang, X., Lu, J., & Hu, J.** (2017). An ensemble deep convolutional neural network model with improved DS evidence fusion for bearing fault diagnosis. *Sensors*, *17*(8): 1729.

[47] **Pan, J., Zi, Y., Chen, J., Zhou, Z., & Wang, B.** (2017). LiftingNet: A novel deep learning network with layerwise feature learning from noisy mechanical data for fault classification. *IEEE transactions on industrial electronics*, *65*(6): 4973-4982.

[48] **Guo, S., Yang, T., Gao, W., Zhang, C., & Zhang, Y.** (2018). An intelligent fault diagnosis method for bearings with variable rotating speed based on pythagorean spatial pyramid pooling CNN. *Sensors*, *18*(11): 3857.

[49] **Qian, W., Li, S., Wang, J., An, Z., & Jiang, X.** (2018). An intelligent fault diagnosis framework for raw vibration signals: adaptive overlapping convolutional neural network. *Measurement Science and Technology*, *29*(9): 095009.

[50] **Chen, Y., Peng, G., Xie, C., Zhang, W., Li, C., & Liu, S.** (2018). ACDIN: Bridging the gap between artificial and real bearing damages for bearing fault diagnosis. *Neurocomputing*, *294*: 61-71.

[51] **Guo, L., Lei, Y., Li, N., & Xing, S.** (2017). Deep convolution feature learning for health indicator construction of bearings. *2017 Prognostics and System Health Management Conference (PHM-Harbin)*.

[52] **Ding, X., & He, Q.** (2017). Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, *66*(8): 1926-1935.

[53] **Jia, F., Lei, Y., Lin, J., Zhou, X., & Lu, N.** (2016). Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing*, *72*: 303-315.

[54] **Mao, W., He, J., Li, Y., & Yan, Y.** (2017). Bearing fault diagnosis with auto-encoder extreme learning machine: A comparative study. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *231*(8): 1560-1578.

[55] **Pan, H., He, X., Tang, S., & Meng, F.** (2018). An Improved Bearing Fault Diagnosis Method using One-Dimensional CNN and LSTM. *Strojniski Vestnik/Journal of Mechanical Engineering*, *64*.

[56] **Jiang, H., Li, X., Shao, H., & Zhao, K.** (2018). Intelligent fault diagnosis of rolling bearings using an improved deep recurrent neural network. *Measurement Science and Technology*, *29*(6): 065107.

[57] **Malik, N.** (2005). Artificial Neural Networks and their applications. *arXiv preprint cs/0505019*.

[58] **Schmidhuber, J.** (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*: 85-117.

[59] **Rosenblatt, F.** (1961). *Principles of neurodynamics. perceptrons and the theory of brain mechanisms*.

[60] **Rumelhart, D. E., Hinton, G. E., & Williams, R. J.** (1986). Learning representations by back-propagating errors. *nature*, *323*(6088): 533-536.

[61] **Hagan, M. T., & Menhaj, M. B.** (1994). Training feedforward networks with the Marquardt algorithm. *IEEE transactions on Neural Networks*, *5*(6): 989-993.

[62] **Hagan, M. T., Demuth, H. B., & Beale, M.** (1997). Neural network design. PWS Publishing Co.

[63] **Møller, M. F.** (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural networks*, *6*(4): 525-533.

[64] **Nocedal, J.** (1980). Updating quasi-Newton matrices with limited storage. *Mathematics of computation*, *35*(151): 773-782.

[65] **Liu, D. C., & Nocedal, J.** (1989). On the limited memory BFGS method for large scale optimization. *Mathematical programming*, *45*(1): 503-528.

[66] **Wolfe, P.** (1969). Convergence conditions for ascent methods. *SIAM review*, *11*(2): 226-235.

[67] **Wolfe, P.** (1971). Convergence conditions for ascent methods. II: Some corrections. *SIAM review*, *13*(2): 185-188.

[68] **Mallat, S. G.** (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, *11*(7): 674-693.

[69] **Bayram, D., & Şeker, S.** (2013). Wavelet based Neuro-Detector for low frequencies of vibration signals in electric motors. *Applied Soft Computing*, *13*(5): 2683-2691.

[70] **Bayram, D., & Şeker, S.** (2014). Wavelet based trend analysis for monitoring and fault detection in induction motors. *2014 IEEE 15th Workshop on Control and Modeling for Power Electronics (COMPEL)*.

[71] **Abbate, A., DeCusatis, C., & Das, P. K.** (2012). Wavelets and subbands: fundamentals and applications. Springer Science & Business Media.

[72] **Bayram, D.** (2015). Condition monitoring and fault detection for induction motors by spectral trending and stationary wavelet analysis. (PhD Thesis). İTÜ Fen Bilimleri Enstitüsü, ISTANBUL.

[73] **Ayaz, E., Ozturk, A., & Seker, S.** (2006). Continuous wavelet transform for bearing damage detection in electric motors. *MELECON 2006-2006 IEEE Mediterranean Electrotechnical Conference*.

[74] **Rad, M. K., Torabizadeh, M., & Noshadi, A.** (2011). Artificial Neural Network-based fault diagnostics of an electric motor using vibration monitoring. *Proceedings 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE)*.

[75] **Ghorban Zadeh Badeli, M., & Bayram, D.** (2021). Feed forward neural network based classification for electric motors. 2021 ISPEC 10th International Conference on Engineering and Natural Sciences.

[76] **Hoang, D.-T., & Kang, H.-J.** (2019). A survey on deep learning based bearing fault diagnosis. *Neurocomputing*, *335*: 327-335.

[77] **Shannon, C.** (1948). A mathematical theory of communication. The Bell SystemTechnical Journal, 27: 379-423. In (pp. 29-115): University of Illinois Press Urbana, IL.

[78] **Lu, S.-D., Sian, H.-W., Wang, M.-H., & Liao, R.-M.** (2019). Application of extension neural network with discrete wavelet transform and Parseval's theorem for power quality analysis. *Applied Sciences*, *9*(11): 2228.

CURRICULUM VITAE

Name Surname : Mina GHORBAN ZADEH BADELI

EDUCATION :

- **B.Sc.** : 2014, Qazvin University, Electronics

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Ghorban Zadeh Badeli, M., & Bayram, D.** (2021). Feed forward neural network based classification for electric motors. 2021 ISPEC 10th International Conference on Engineering and Natural Sciences.
- **Ghorban Zadeh Badeli, M., & Bayram, D.** (2021). Back propagation neural network based classification for electric motors. In Ş. Yücelbaş (Ed.), *Machine learning, energy and industrial applications in technology and engineering sciences* (pp. 51-80). İksad Publishing House.