ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

MSc THESIS

Jean de Dieu UWISENGEYIMANA

DIAGNOSING KNEE PATHOLOGY USING SURFACE ELECTROMYO-GRAPHY (SEMG) AND ARTIFICIAL NEURAL NETWORKS

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

ADANA-2017

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

DIAGNOSING KNEE PATHOLOGY USING SURFACE ELECTROMYOGRAPHY (SEMG) AND ARTIFICIAL NEURAL NETWORKS

Jean de Dieu UWISENGEYIMANA

MSc THESIS

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

We certify that the thesis titled above was reviewed and approved for the award of degree of the Master of Science by the board of jury on 27/09/2017.

…….....	…….....	.....
Assoc. Prof. Dr. Turgay İBRİKÇI	Assoc. Prof. Dr. Sami ARICA	Assist. Prof. Dr.İrem Ersöz KAYA
SUPERVISOR	MEMBER	MEMBER

This MSc Thesis is written at the Department of Institute of Natural And Applied Sciences of Çukurova University.
Registration Number:

Prof. Dr. Mustafa GÖK
Director
Institute of Natural and Applied Sciences

This study was supported by Research Projects Unit BAP Coordination Unit of Çukurova University (Project Number : FYL-2017-7864) under the fellowship with TUBITAK-BIDEB 2235.

Not: The usage of the presented specific declerations, tables, figures, and photographs either in this thesis or in any other reference without citation is subject to "The law of Arts and Intellectual Products" number of 5846 of Turkish Republic

ABSTRACT

MSc THESIS

DIAGNOSING KNEE PATHOLOGY USING SURFACE ELECTROMYOGRAPHY (SEMG) AND ARTIFICIAL NEURAL NETWORKS

Jean de Dieu UWISENGEYIMANA

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF ELECTRICAL AND ELECTRONICS NGINEERING

Supervisor : Assoc. Prof. Dr. Turgay IBRIKCI
Year: 2017, Page: 82
Jury : Assoc. Prof. Dr. Turgay İBRIKÇI
: Assoc. Prof. Dr. Sami ARICA
: Asst. Prof. Dr. İrem Ersöz KAYA

Among various medical diagnostic tests performed to identify pathology in the knee, especially knee osteoarthritis, most of them are invasive and expensive. In this thesis, another methodology for diagnosing the abnormality in the knee in a more quick, non-invasive and cheap manner has been investigated. For that purpose, surface Electromyography (SEMG) signals recorded from the four muscles surrounding the knee, the recording of the flexion degree in the knee and pattern recognition algorithms were used. The datasets of this experiment comprised 22 subjects among whom 11 subjects had normal knee and other 11 subjects had an osteoarthritis-affected knee. Time-series features were then extracted from each frame and were used to train, test and validated 13 different learning classifiers and 7 variants of deep neural networks. In this study, the best performance measure above 99% was achieved by most of the tested classifiers under the group of Support Vector Machines (SVM), multilayer Perceptron (MLP), k-Nearest Neighbor (K-NN) and Ensemble classifiers. Whereas, the use of deep nets which were investigated over a wide range of hidden size of the autoencoders, showed accuracy of 86.6% with final softmax layer and accuracy of 91.3% by replacing the final softmax layer with k-nearest neighbor. By comparison, artificial neural networks outperformed deep neural networks and it is therefore concluded that the knee pathology can be diagnosed more efficiently and automatically using surface electromyography signals and artificial neural network algorithms.

Keywords: EMG, Osteoarthritis, Feature Extraction, Neural Networks, Deep Networks.

ÖZ

YÜKSEK LİSANS TEZİ

YÜZEY ELEKTROMİYOGRAFİSİ (SEMG) VE YAPAY SİNİR AĞLARI KULLANARAK DİZ PATOLOJİSİNİN TEŞHİSİ

Jean de Dieu UWISENGEYIMANA

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANA BİLİM DALI

Danışman : Doç. Dr. Turgay İBRIKÇİ
Yıl: 2017, Sayfa: 82
Jüri : Doç. Dr. Turgay İBRIKÇİ
: Doç. Dr Sami ARICA
: Yrd. Doç. Dr. İrem Ersöz KAYA

Diz, özellikle diz osteoartritinde patolojiyi saptamak için yapılan çeşitli tıbbi tanı testlerin çoğu invaziv ve pahalıdır. Bu tezde, dizdeki anormalliklerin daha hızlı, invaziv olmayan ve ucuz şekilde teşhis edilmesi için diğer yöntemlerin araştırılmasıdır. Bu amaçla, diz çevresindeki dört farklı kastan kaydedilen yüzey Elektromiyografi (SEMG) sinyalleri, dizin bükülme derecesi ve tanıma algoritmaları kullanıldı. Bu veri setinde, 11 denekte normal diz ve diğer 11 denekte ise osteoartritten etkilenen diz bilgilerinden oluşantoplamda 22 denekten oluşmaktadır. Bu veri setinde her pencerede zaman serisine bağlı özellikleri çıkarıldı. Bu özellikleri kullanan 13 farklı öğrenme sınıflandırıcısı ve 7 derin sinir ağları ile eğitilmesi, test edilmesi ve doğrulanması için kullanılmıştır. Bu çalışmada,%99'u aşan en iyi performans çok katmanlı algılayıcı olarak Destek Vektör Makineleri (SVM), vetopluluk sınıflandırıcıları grubunda ise k-en yakın komşusu (k-NN) ile başarılmıştır. Buna karşın, otoencoderlerde gizli büyüklükte geniş bir yelpazede araştırılan derin ağların kullanımı son softmax katmanı ile%86.6 doğruluk göstermiştir ve son softmax katmanı k-en yakın komşusu (k-NN) ile değiştirilerek doğrulukta %91.3'ye ulaşılmıştır. Karşılaştırma sonuçlarında, yapay sinir ağları derin sinir ağlarından daha üstün performans göstermiştir. Bu nedenle, diz patolojisinin yüzeyelektromiyografi sinyalleri kullanılarak yapay sinir ağı algoritmaları ile daha verimli ve otomatik teşhis edilebileceği sonucuna varılmıştır.

Anahtar kelimeler: EMG, Osteoartrit, Özellik çıkarma, Sinir ağları, Derin ağlar.

EXTENDED ABSTRACT

In this thesis, the main purpose was to investigate an Artificial Neural Network (ANN)-based clinical decision support system that is intented to help doctors predict whether a visiting patient is suffering from knee osteoarthritis or has got any other knee abnormality. The current one way to detect osteoarthritis is to inspect the knee images taken by means of X-Ray and to computerize the identification of the knee pathologies using image classification techniques. Indeed, the diagnosis using X-Ray tests would be powerful and adequate means to identify any damages in bones and joints, but, this way is not profitable and for this reason it not most often adopted the earliest possible like the inexpensive ways. Another way stated in the literatures was to employ Magnetic Resonance Imaging (MRI), which uses a magnetic field to shoot many images of the knee and is capable of executing pictures of the cartilage and the ligaments. The current researches discuss the fresh beginning of employing MRI to get clues and reports about the status of the cartilage, bones and the joints. Neverthless, diagnosis using MRI tests are not profitable as well and many clinics and hospitals located in Africa and Asia do not own these expensive medical equipments. For cheap and quick detection of the problem, it is proposed to use Computer- Aided Diagnosis systems.

In fact, artificial neural networks are among the current Computer-Aided Dignosis (CAD) applications employed to build clinical decision support system, because they are less complex algorithms in architecture, data structure and user interface. The main reasons that motivated the utilization of computers in diagnosis tasks are the difficulties that doctors meet while making medical diagnosis by their own. For example; to make the best diagnostic decision, the physician needs sufficient number of experienced cases of the disease and even the most experienced doctors have difficulties diagnosing rare or new diseases. Also, sentimental issues and body weakness corrupt performance of the physician. So,

following these problems, a computer system could address them as it never gets annoyed, fatigued or exhausted, can be upgraded in an eye blinking time, and far better it is not expensive and can be shared. It can also work alone for some cases and hence make medical specialists free to concentrate on big and more severe cases.

To realize the new methodology of this thesis to diagnose the knee pathology, Surface Electromyography (SEMG) signals recorded from the four muscles surrounding the knee, the recording of the flexion degree in the knee and pattern recognition algorithms were used. The datasets of this experiment comprised 22 subjects among whom 11 subjects had normal knee and other 11 subjects had an osteoarthritis-affected knee.

After having surface EMG, the signal was filtered using a 20th order Chebyshev type II band-pass filter. The bandpass filter was used to maintain the useful signal contained within the signal bandwidth by rejecting the signals whose frequencies are outside the useful frequency range (low and high frequency signals). The low frequency cutoff of bandpass filter overthrows baseline drift which is from time to time caused by slight migration of electrodes, sweating of the skin. It also suppresses any DC offset. Typical values for the low frequency cutoff range from 5 to 20 Hz. High frequency cutoff of band pass filter gets rid of high frequency noise and hinders aliasing from taking place in the sampled signal. The high frequency cutoff should be sufficiently high in order to clearly keep capturing the qwick on-off bursts of the EMG signal. Regular values for high frequency cutoff are in the range of 200Hz–1kHz. In this thesis, the designed filter was used to remove signals that are outside the extent of the electromyogram between 20Hz and 480Hz. The filter's attenuation of unwanted signals or amplitude stop (Astop) was 40dB.

In this thesis, to reduce the dimensionality of SEMG data, the data were segmented using a chain of time-overlapping analysis windows. The initial total sample size was 1,048,576 samples which were pre-processed and segmented

using 90% overlaping-windows of 5000 samples each. The result of this technique was a resampled dataset with 2642 samples.

After signal segmentation, time series features extraction was performed. In order to capture the most important knowledge or information in the electromyogram, feature extraction procedure was performed. The reaped features of the electromyogram are usually computed depending on linear or statistical methods. As seen in many literatures, a number of significantly extracted features depend on time-domain statistics such as the mean of the samples, the stand deviation of the samples, the variance of the sample signals, the zero crossing value, slope sign changes, which need less computational resources. The time-domain methods have abundantly been also employed in many previous researches for feature extraction, due to the reason that they comparatively present no difficulties in their implementation and present high performance. More specificallt, in this thesis time- series features such as mean power of the original signals, mean frequency of the original signals, peak value of the original signals, mean of the original signals, standard deviation of original signals and variance of the original signal were extracted from each signal frame and were used to train, test and validated 13 different learning classifiers namely Complex Tree (CT), Medium Tree (MT), Linear Discriminant Analysis (LDA), Logistic regression (LR), MultiLayer Percetron (MLP), BayesNet, Linear SVM (LSVM), Quadratic SVM (QSVM), Bagged Trees, Boosted Tree, Fine K-NN, Medium K-NN,Coarse K-NN and 7 variants of deep neural network.

In this study, the best accuracy above 99% was reached by the most of the investigated algorithms under the group of Support Vector Machines (SVM), MultiLayer Perceptron (MLP), k-Nearest Neighbor (K-NN) and Ensemble classifiers. In comparison, the possibility of validating the use of deep learning to solve the same problem was no better. The accuracy achieved by deep learning was 86.6% with softmax algorithm in the final layer and a performance measure of 91.3% by substituting the softmax algorithm with with k-nearest neighbour in the

final layer. By comparison, artificial neural networks outperformed deep neural networks.

In conclusion, we achieved all the set-out objectives and we showed that it is possible and of great importance to predict any knee pathology using the data recorded by SEMG and by goniometer, then classify them using Artificial Neural Network (ANN).

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor, Associate Prof. Dr Turgay IBRİKCİ, who benefited me with his expertise, patience, guidance and insightful criticisms throughout the learning process of this master's degree program. I appreciate his vast knowledge in many research areas like Computational Biology, Neural Networks, Pattern Recognition and Software Engineering, which were needed for the completion of this thesis. I also thank him for assisting me in writing many academic publications and this thesis report as well.

Special thanks also go out to Associate Prof. Dr Ahmet TEKE, who provided me with directions, technical support and became a friend, than a professor. It was though his persistence, understanding and kindness that I completed my second manuscript publication. I doubt that I will ever be able to convey my appreciation fully to him, but I owe him my eternal gratitude.

I must also acknowledge Associate Prof. Dr. Sami ARICA for his suggestions, extra hours and practical instructions about biomedical signal processing techniques. My appreciations also go out to Assistant Prof. Dr. Murat AKSOY and Assistant Prof Mehmet Uğraş CUMA for their helpful lectures and lab practices.

Finally, I must recognize that this research would not have been possible without the financial assistance of Cukurova University Project Unit (Project Number: FYL-2017-7864), The Scientific and Technological Research Council of Turkey (TUBITAK-BIDEB 2235) and the University of Rwanda (UR). I sincerely express my gratitude to those agencies.

CONTENTS **PAGE**

ABSTRACT I
ÖZ II
EXTENDED ABSTRACT III
ACKNOWLEDGEMENTS VII
CONTENTS VIII
LIST OF TABLES X
LIST OF FIGURES XII
LIST OF ABBREVIATIONS XIV
1. INTRODUCTION 1
1.1. Importance of the Study. 3
1.2. Objectives 5
2. REVIEW 7
2.1. Artificial Neural Networks in Biomedicine 7
2.1.1. Overview of an Artificial Neural Networks 7
2.1.2. Applications of Artificial Neural Networks İn Biomedicine . 10
2.2. Electromyography and its Applications in Biomedecine 11
2.2.1. Overview of Electromyography 12
2.2.2. Applications of Electromyography in Medical Diagnosis 13
2.3. Human Knee Structure and Knee Muscles 14
2.4. Biomedical Signal Processing Techniques. 16
3. MATERIALS AND METHODS 19
3.1. Discription of Lower Limb EMG Datasets. 19
3.2. Data pre-processing 20
3.2.1. Normalization 20

3.2.2. EMG Signal Processing and Feature Extraction. 22
3.3. Classification Algorithms 23
3.3.1. Multilayer Perceptron (MLP) 23
3.3.2. Support Vector Machine (SVM) 26
3.3.4. Bayesian Networks 30
3.3.5. Complex and Medium Decision Trees 32
3.3.6. K-Nearest Neighbors 34
3.3.7. Linear Discriminant Analysis 35
3.3.8. Ensemble Classifiers. 35
3.3.9. Deep Neural Network 38
4. RESULTS AND DISCUSSIONS 41
4.1. Analysis Results of EMG Signal 41
4.2. EMG Signal Conditioning: Filtering and Noise Reduction 45
4.3. Results of EMG Feature Extraction 46
4.4. Results of classification algorithms 49
5. CONCLUSION 59
REFERENCES 61
CURRICULUM VITAE 79
APPENDICES 80

LIST OF TABLES **PAGE**

Table 4.1. Performance results from the 13 tested learning classifiers 51
Table 4.2. Results from new methods attempted to improve the performance of the deep learning model..... 57

LIST OF FIGURES **PAGE**

Figure 1.1. Main pertinence areas of artificial neural networks in medicine 4
Figure 2.1. Relationship between the natural neuron and the artificial neural network 9
Figure 2.2. Waveform of a typical surface EMG signal.....13
Figure 2.3. The structure of the knee.....15
Figure 2.4. Muscles of the Lower Limb: The rear and back views. 16
Figure 2.5. Stages of signal processing for biomedical application 17
Figure 3.1. Stages of an electromyographic pattern classification system.....19
Figure 3.2. Block diagram of data preprocessing.....20
Figure 3.3. Originally unnormalized data and normalized data using Zscore normalization.....21
Figure 3.4. Originally unnormalized data and normalized data using sigmoidal ..22
Figure 3.5. Block diagram of a multilayer perceptron with two hidden layers 24
Figure 3.6. Deciding the optimal hyperplane 26
Figure 3.7. Form of the logistic function (Mitchell, 2015).....30
Figure 3.8. Illustrating the components of the binary decision tree 32
Figure 3.9. Mapping data from 4 dimensions to 2 dimensions using a ANN with one hidden layer: a, shows a linear autoencoder and b, shows a non-linear autoencoder.....39
Figure 4.1. Plots of signals in all the four different muscles of interest for the abnormal patient in marching position 42
Figure 4.2. Plots of signals in all the four different muscles of interest for the normal patient in marching position.....42
Figure 4.3. Plots of signals in all the four different muscles of interest for the normal patient in seated position.....43
Figure 4.4. Plots of signals in all the four different muscles of interest for the abnormal.....43

Figure 4.5. Plots of signals in all the four different muscles of interest for the normal patient in standing position 44
Figure 4.6. Plots of signals in all the four different muscles of interest for the abnormal patient in standing position 44
Figure 4.7. Magnitude Response of Chebyshev type II filter 46
Figure 4.8. Phase Response of Chebyshev type II filter 46
Figure 4.9. Segmentation of analysis windows of EMG recordings 47
Figure 4.10. Comparing performance metrics among different groups of learning classifiers. 53
Figure 4.11. Comparing performance metrics of ensemble and K-NN classifier groups 54
Figure 4.12. The architecture of the deep learning system 55
Figure 4.13. Results of the achieved accuracy by varying the number of neurons in the two autoencoders 55
Figure 4.14. Different performance measure achieved by adjusting maximum epoch parameter of the DNN 56
Figure 4.15. Performance comparison among the new methods of DNN 58

LIST OF ABBREVIATIONS

ACL	: Anterior Cruciate Ligament
ANNs	: Artificial Neural Networks
AUC	: Area Under ROC Curve
BF	: Biceps Femoris
BNs	: Bayesian Networks
CAD	: Computer-Aided Diagnosing
CT	: Complex Tree
DBN	: Deep Belief Networks
DC	: Direct Current
DNN	: Deep Neural Network
DTs	: Decision Trees
ECG	: Electrocardiogram
EEG	: Electroencephalogram
EMG	: Electromyogram
EOG	: Electrooculogram
FFT	: Fast Fourier Transform
FN	: False Negative.
FP	: False Positive
GMs	: Graphical Models
K-NN	: k-Nearest Neighbors
LDA	: Linear Discriminant Analysis
LR	: Logistic Regression
LSVMs	: Linear Support Vector Machines
MLP	: Multilayer Perceptron
MRI	: Magnetic Resonance Imaging
MT	: Medium Tree
OA	: Osteoarthritis

QSVMs : Quadratic Support Vector Machines

RF : Rectus Femoris

ROC : Receiver Operating Characteristic Curve

SEMG : Surface Electromyography

ST : Semitendinosus

SVM : Support Vector Machines

TN : True Negative

TP : True Positive

VM : Vastus Medialis

1. INTRODUCTION

One of the bone joint pathologies is Osteoarthritis (OA), which affects both load bearing and non-weight bearing cartilages (Gregson et al., 2017), (Kean et al., 2017). In fact, a healthy cartilage acts as a cushion which allows joints of bones to slide agains one another so as to resist any damage to the joint. By contrast, according to (Anjani and Kubakaddi, 2014) and (Kanneppady and Kanneppady, 2015), with an osteoarthritis condition, the upper part of the cartilage fails and decays and this causes inflammation and soreness in the bone joints. For years, osteoarthritis was thought of to be a result of normal wear and tear of cartilage during a person's lifetime; on the contrary, researchers nowadays recognize that several factors may lead to the development of this condition including any damage due to joint injury, such as a fracture or torn on cartilage, infection in the joint or bone, an occupation involving heavy physical work, being overweight over a long period of time (Alnahdi et al., 2012).

As reported by several studies, the knee joint is the most part of the body which is susceptible to damage because it supports the heaviest burden of the personal chassis as it promotes a smooth movement of the body. (Janidarmian et al., 2014), (Chen et al., 2015). Severe knee shocks, like those that affect the Anterior Cruciate Ligament (ACL), are a major motive for study involvement. These injuries to ACL can promote failure for the joint to digest the shock and this may result in much more pressures across the knee. These pressures therefore becomes more intensified, which can prompt a series of knee damages, meniscus impairment, and poor alignment (Robert et al., 2008). The highest occurrence is found in young people who play games like football, basketball, handball, etc... (Bahr et al., 2005). These shocks may cause grave effects for the injured player, in terms of not only treatment expenses and many missed matches, but also a pronounced inclinement to early osteoarthrosis (McAlindon et al., 2014), (Teitel and Zieve, 2010). For all these reasons, it could therefore be of great benefit to

devise easy, inexpensive and qwick means to test any abnormality of the knee for the sake of early detection and treatment of the problem.

Many researches have been conducted to contribute to the effective way for detection of osteoarthritis. The studies of Shamir et al. (2009), showed that one way was to inspect the knee images taken by means of X-Ray and to computerize the identification the knee pathologies using image classification techniques. Indeed, the diagnosis using X-Ray tests would be an powerful and adequate means to identify any damages in bones and joints, but according to (Bedson et al., 2003), this way is not profitable and for this reason it not most often adopted the earliest possible like the inexpensive ways. Another way stated in the literatures was to employ Magnetic Resonance Imaging (MRI), which uses a magnetic field to shoot many images of the knee and is capable of executing pictures of the cartilage and the ligaments. Studies done by (Eckstein et al., 2006) and (Peterfy and Kothari, 2006), discussed the fresh beginning of employing MRI to get clues and reports about the status of the cartilage, bones and the joints. Neverthless, diagnosis using MRI tests are not profitable as well and many clinics and hospitals located in Africa and Asia do not own these expensive medical equipments.

In addition, the study conducted by Gonzalez et al. (2015), showed that the status of quadriceps muscles might give full information about abnormal instability of the ligaments in the knee, or can show if there are any other unperceived damages in the knee. In their study, the important muscles to investigate were shown up to be the rectus femoris, the vastus medialis, biceps femoris and Semitendinosus.

This study will consist of building a Neural Network model which is proposed as a new and cheap method for diagnosing knee abnormalities. An investigation of of several learning classifiers comprising ANN and deep learning networks was made to achieve the best performance. The network will be trained by the processed signals recorded by the Surface Electromyography (SEMG) and

the goniometer. Design of digital filters and Feature extraction algorithms will be of crucial importance prior to digital classification of the signals.

1.1. Importance of the Study.

Since the early 1960s, the computing community has made efforts regarding use of computer-assisted techniques for some biomedical applications (Shortliffe and Cimino, 2014). These methods have served well in keeping medical records in digital form. On the other hand, as pointed out by Szolovits et al., (1988), in case clinical examination, due to the complication of the duty, it has not been easy or executable yet to hope for a completely robotized or computerized medical examining model. The reason for this was due to the fact that biomedical systems have some components that are well defined, along with numerous others that are only partially understood.

However, recent development in learning machines like Artificial Neural Networks (ANNs) have been a 'hot' research area and it is expected that they will receive considerable adoption to biomedical diagnostic systems in the coming years (karamzadeh et al., 2013). Nowadays, artificial neural networks are being used to promote potent and important devices to benefit medical professionals to analyze, model and make sense of complicated clinical records across a wide range of medical applications shown in Figure 1.1.

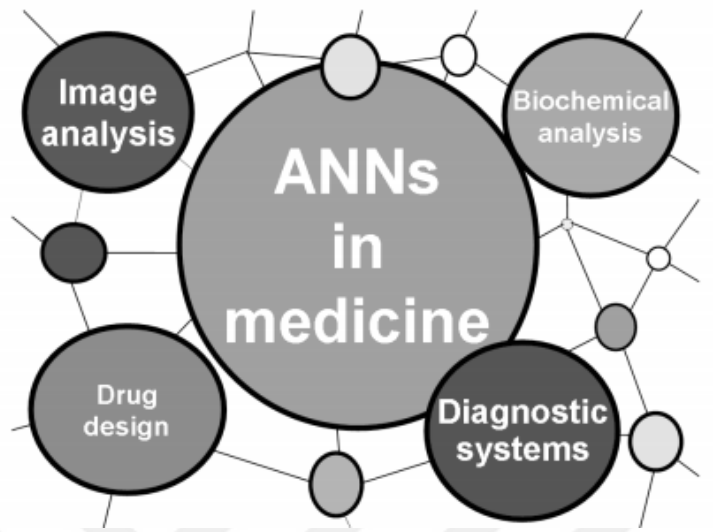


Figure 1.1. Main pertinence areas of artificial neural networks in medicine (Amato et al., 2013)

In most studies, pertinence of Artificial Neural Networks (ANNs) in medical field are supervised learning problems (Kumar, 2012). In all these studies, ANNs are employed as a powerful discriminating classifier.

The main reasons that motivated the utilization of computers in diagnosis tasks are the difficulties that physicians face while making medical diagnosis by their own. Among those difficulties we cite:

- To make the best diagnostic decision, the physician needs sufficient number of experienced cases of the disease.
- Even the most experienced doctors have difficulties diagnosing rare or new diseases.
- sentimental issues and body weakness corrupt performance of the physician.
- The training policy for medical specialists takes a long time and very expensive.

- A number of people who go to a health center or a hospital might not be seriously sick and the doctors could waste their time on them.

So, following these problems, a computer system can address them as it never gets annoyed, fatigued or exhausted, can be upgraded in an eye blinking time, and far better it is not expensive and can be shared. It can also work alone for some cases and hence make medical specialists free to concentrate on big and more severe cases.

1.2. Objectives

The main objective of this thesis is to build a sytem to automatize the diagnosis process or to aid doctors to detection of any abnormality in the knee joint.

Other objectives are:

- Pre-processing surface EMG and goniometry raw data with time overlapping analysis windows.
- Extracting time domain features from each analysis window prior to classification task.
- Training, testing and validating different learning classifiers and deep learningalgorithms.

2. REVIEW

In recent decades, computational methods have been widely used in the biomedicine researches, including medical image analysis, diagnosing systems, and detection of cancerous tumors. In particular, many computational problems can be formulated as the prediction task on the biomedical data, such as tumor images generated by Magnetic Resonance Imaging (MRI), electronic medical records using biomedical sensors such as Electroencephalogram (EEG), Electrocardiogram (ECG), Electromyogram (EMG) or Electrooculogram (EOG). In many researches, one mostly used computational methods to extract knowledge from biomedical data are Artificial Neural Networks (ANNs) (Hapudeniya, 2010), (Lisboa et al., 2010), (Chen et al., 2005).

In the following sub-sections, we firstly review a brief history of artificial neural networks and its medical diagnostic applications, Application of ANNs on EMG datasets, a qwick review on the knee structure and knee muscles and we finally review the biomedical signal processing techniques.

2.1. Artificial Neural Networks in Biomedicine

Since a long time ago, even before the appearance of Artificial neural Networks (ANNs) concept, researchers had sought to use computers to support clinical decisions made by medical professionals. The following discussion is about the brief information about ANNs, its history and its applications in medical field.

2.1.1. Overview of an Artificial Neural Networks

The first study on Artificial Neural Networks (ANNs) was a result of intensive researches carried out by McCulloch and Pitt (1943). Afterwards, in 1958, the computer and the brain were related and reported about which enabled further studies on human nervous system and the building systems based on

artificial neural networks, which imitate the way humankind deals with sensory information (Birdi et al., 2013), (Furber and Temple, 2007).

The first examination of ANNs developed in 1940s and initially referred to as the perceptron, did not show satisfactory results and presented several shortcomings such as dealing only with linearly correlated data. However, due to these limitations of the perceptron, researchers sought to extend the initial perceptron architecture, which was previously a single-layer neural system to a multilayer neural systems (Zupan and Gasteiger, 1991), (Smits et al., 1994). The new multilayer perceptron architecture was first described by Hopfield (1982) with the introduction of how to handle non-linearity between training samples and response and this new approach resulted in a more enhanced ANNs. In addition to Holpfield's study, Werbos (1974) had proposed the back-propagation learning algorithm and the combination of the two contributed to the popularity and universality of ANNs.

Actually, an ANN is an architecture to easily handle or process information in a parallel way. The system is consisted of a big number of simple processing units called artificial neurons also called nodes that are hooked by direct connections, and which work together to solve concurrent operations so as to find a solution of a given problem (Haykin, 1994). An important trait of ANNs is their capability to copy with changes in the environment by altering their connection weight or structure (Pratap and Shelja, 2013). This particularity makes it easy for neural network systems to learn automatically.

In general, ANN procedures are a part of mathematical models that are based on the way the human brain works (Maltarollo et al., 2013). What all ANN techniques have in common is the concept of "neurons". This term "neuron" implies that there is close similarity between artificial neuron and the biological one, as is shown in Figure 2.1. According the reports of (Perwej and Parwej, 2012), (Lakra et al., 2012), the human brain has approximately 100 billion of neurons. Where, every neuron is associated to thousands of other neurons and they all

together make a biological neural network which commute information in the form of stimuli, in and out of the human brain consistently.

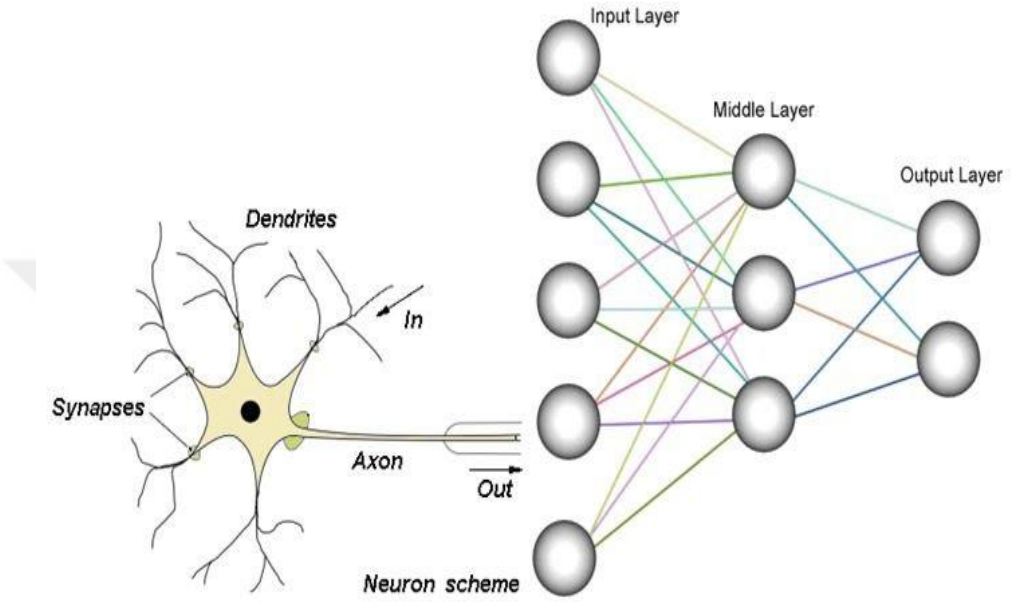


Figure 2.1. Relationship between the natural neuron and the artificial neural network (Barber, 2007)

Each neuron in the system is composed of 4 major parts: the synapses, soma, axon, and dendrites. The soma is the body of the cell which holds particlar organelles to furnish with energy and process information. In the nervous system, each neuron accepts information in the form of electrical signals which come from nearby neurons and passing through out a huge number synapses (Stiles and Jernigan, 2010). When sufficient of this charge is acquired, the neuron is switched on to send out a signal to be collected up by the following neuron's dendrites.

Similarly, the main concept in the wake of artificial neural networks started from the well-known theory of the psychologist Hebbs (1949) which states: “When an axon of cell A is close enough to trigger cell B or frequently or diligently

involves in ignite it, some growth process or metabolic change occur in one or both cells such that the readiness of cell A, as one of the cells igniting cell B, is marked up". This statement is met in ANNs learning rules, which shows that the theory of neurons is wholly applied to artificial neural networks. The axons are the links between neurons and these links are called weights in terms of ANN. The cability of a neuron to fire when many other neurons attempt to excite is compared to something called activation-function within the ANN terminologies.

2.1.2. Applications of Artificial Neural Networks in Biomedicine

Ever since 1950s it was known that computers could be used for disease diagnosis. Although many attempts were made in that decade on designing Computer-Aided Diagnosing (CAD) systems, those attempts were not successful according to Engle (1992), who after several years of experiments stated an outcome saying that they could not have much hope in the evolution of computer softwares to stand for the role of a doctor for decision making.

Despite this conclusion, researchers continued and in 1960s first findings about computable analysis of therapeutical images using a computer were reported by many researches such as (Lodwick et al., 1963), (Myers et al., 1964), (Winsbarg et al., 1967). In this decade, it was generally supposed that machines could stand in the placeof physicians in diagnosing pathologies, because computers and machines could perform some tasks better than are people do. Still, all the attempts made were not successful, due to reason that machines were not satisfactorily powerful and advanced, signal processing techniques were not available yet, and medical data were not easily accessible.

In the next decade, many other researchers including De Dombal et al. (1972), Kruger et al. (1972), Toriwaki et al. (1973), and again Kruger et al. (1974), kept to work in the area of the computational medical decision. A study of interest in this decade was reported by De Dembal who wrote a paper comparing the accuracy of the physician and the computer in diagnosing a course of three hundred

and four people having abdominal pain of sensitive outbreak. The computer's complete detection efficiency of around 91.8% was much higher than that of the most experienced specialists of the clinical team consulting one another with accuracy of 79 6%. Following these studies, it was then offered that the use of such a models to help the clinician was both possible in a ordinary clinical environment, and could apparently be of reasonable benefits. Furthermore, within the same decade, there were other researches that focused on the diagnosis of acute abdominal pain such as Wilson et al. (1977), Gunn (1975), and Wilson et al. (1975).

In spite of previous findings, important and systematic studies and improvements on different CAD systems were launched in the beginning of the year of 1980 at the Kurt Rossmann laboratories for radiologic image research in the department of radiology at the university of Chicago (Doi, 2007). This was the decade when Sher (1980), associate professor of clinical pathology at New York university school of medicine, developed and reported his research titled "mathematical and computer assisted procedures in clinical decision making." In his findings, he said that in practice, more added data could improve the building of computational models and that with enhancements in computer programs, could be in a good mood to impact the the way forward in medical diagnostic practices in a more inexpensive and profitable ways.

Other numerous reviews regarding the application of ANNs in medical diagnosis were present from 2000 and the a considerable number of studies were interested in chest, breast, colon, brain, liver, skeletal and vascular systems.

2.2. Electromyography and its Applications in Biomedecine

Electromyography(EMG) nowadays has become an important parameter in biomedical and clinical applications. Nevertheless, before applying EMG signals, their detection, processing, interpretation and analysis is needed. EMG are among the bioelectrical signals that fit well as an input for prosthetic device control and

were for long used to develop prosthetic devices that are often used to replace the missing parts of human body. Surface Electromyography (SEMG) has been used successfully for the controlling of manufactured body parts., but its employment to medical diagnostic tests has been very confined or nonexistent (Goen and Tiwari, 2013). Previously, SEMG had problems which limited them from being used as a diagnostic tool (Pullman et al., 2000), (Zwarts et al., 2000). However, electromyogram resulting from voluntary muscle activities, possess more meaningful information than it was previously thought of having (Merletti and De Luca, 1989). In addition, several studies and reports concerning the suitability of SEMG technique for diagnostic purpose are available and are reviewed in this sub-section preceded by an overview of electromyography.

2.2.1. Overview of Electromyography

Electromyography is a technique for evaluating and recording the bioelectric signals resulting from muscle activity. These electrical signals may be detected from the surface of the body close to the muscle of interest (Mankar, 2011). The Electromyography (EMG) was introduced in the first half of the twentieth century, and the first major studies on EMG were performed in the 1940s and 1950s (Steindler, 1953).

To obtain EMG data, two major technologies are basically used by EMG researchers while recording the electromyogram. Those procedures are ingrained EMG and Surface EMG (Williams, 2015). Ingrained EMG depends upon the introduction of an electrode rooted into skeletal muscle, which by its nature is not easy to be used by researchers and confines its employment to clinical laboratories. In contrast, Surface EMG is a non-invasive method that can give computable data on the activity within shallow muscles during its contraction in both laboratory and field environments (Rudroff, 2008), (De Luca, 1997). A typical signal recorded by a surface EMG is shown in Figure 2.2 below.

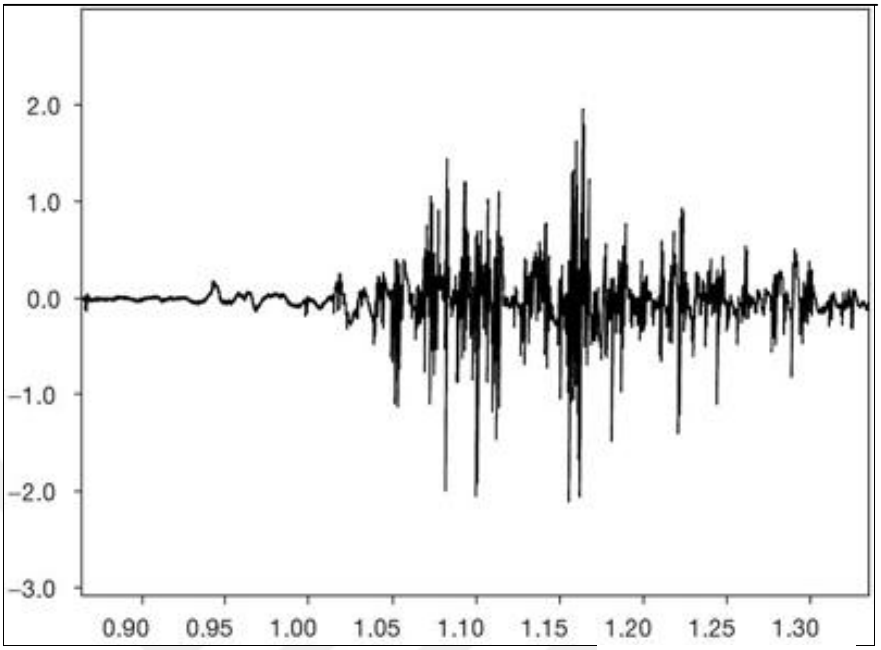


Figure 2.2. Waveform of a typical surface EMG signal (Roberto et al., 2009)

2.2.2. Applications of Electromyography in Medical Diagnosis

In the literatures, there has been a lot of attempts towards using artificial neural networks in analysis and Classification of Electromyogram (EMG) Signals (Ahsan et al., 2012), (Caesarendra et al., 2016), (Liu et al., 2016). Moreover, many researchers used Artificial neural networks with Electromyography (EMG) signals to build automated and robust systems in the field of robotics, rehabilitation, biomedical, etc... (Hargrove et al., 2013), (Kiguchi and Hayashi, 2012) , (Sundaraj, 2009), (Koçer and Tümer, 2016), (Chowdhury et al., 2013) .

However, in this thesis, most of review emphasis has been put on the early researches which used lower limb EMG signals to diagnose knee enjuries. In the study conducted by González et al. (2015), a knee functional state classifier was achieved. The system used a signals resulting from surface Electromyogram and a degree of flexion in the knee, which are processed with biomedical signal processing and analysis procedures in time-frequency space through a spectrogram

and a wavelet transform. MultiLayer Perceptron was used as a learning technique and it showed around 80% of accuracy. However, This accuracy is too low to rely on the system. In this thesis, it sought to improve this accuracy by comparing more machine learning algorithms.

Other several studies reviewed here have focused on assessing the knee Anterior Cruciate Ligament (ACL) injuries by using the electromyography signals and knee flexion angle data (Podraza and White, 2010), (Nagano et al., 2011), (Kernozek and Ragan, 2008) (Henriksen et al., 2009), (Pollard et al., 2010). According to these researches, it is suitable to assess knee status using electromyographic signals and the degree of flexion in the knee.

2.3. Human Knee Structure and Knee Muscles

The knee is the bulkiest bone junction in the human structure. The topmost and lower bones of the knee are detached by two spheres called menisci. The bone of the topmost part of the leg called femur and the lower bones of the leg called tibia and fibula are linked by ligaments, tendons, and muscles. The exterior of the bones innermost of the knee joint is enclosed by articular cartilage, which take up the shock and provides a gentle, floating area for mobility of the joint. Figure 2.3 shows the structure of the knee in detail.

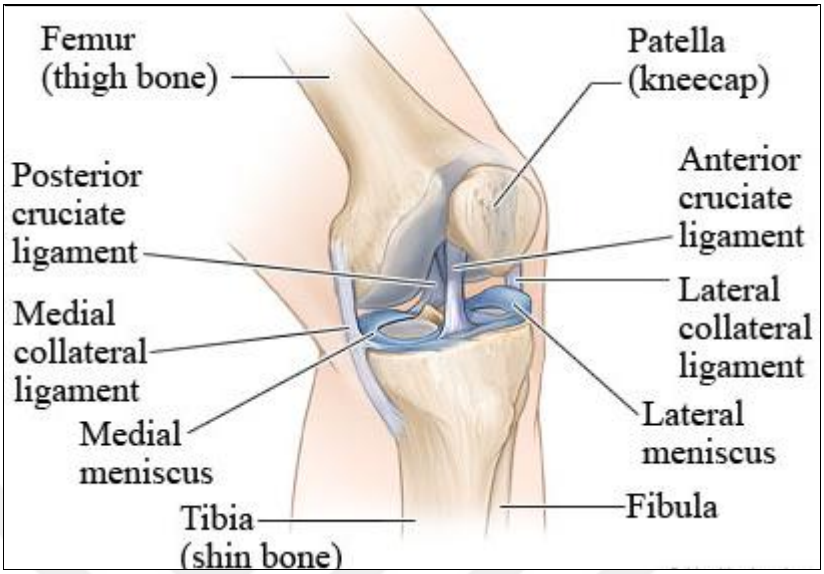


Figure 2.3. The structure of the knee (Webmed, 2017)

Even if knee abnormalities are often due to schocks to one or more of the knee structures, they might be caused by another reasons. Some people have more probability to be affected by knee issues than others. In the studies carried out by Blagojevic et al. (2010), Quintana et al. (2008), Murphy et al. (2008), it was pointed out that old people are the most vulnerable compared to young ones. Among other risks that increase the probability of having a knee complications are gender (Srikanth et al., 2005), obesity (Shiri et al., 2009), (Lementowski et al., 2008), (Toivanen et al., 2009), sports activities (Tran et al., 2016) and occupational activities (McWilliams et al., 2011), or having a disease such as osteoporosis or arthritis increase the chances of having problems with the knees (Woolf and Pfleger, 2003).

According to Bennell et al. (2008), the muscles of the lower limb have a crual advantage in the outset and administration of knee osteoarthritis (OA). In the research conducted by Park et al. (2016), they found that people who have knee abnormality showed decreased hip external rotator, knee extensor, and ankle inversion muscle force output compared with healthy controls, as well as increased

peak knee adduction angles. In addition, studies done by Aslan et al. (2016), Gonzales et al. (2016) point out the relationship between lower limb muscles and knee structure problems.

The lower limb muscles of interest while trying to deal with the knee structural problems or the knee osteorthritis on any abnormalities in the knee, are shown in the Figure 2.4.

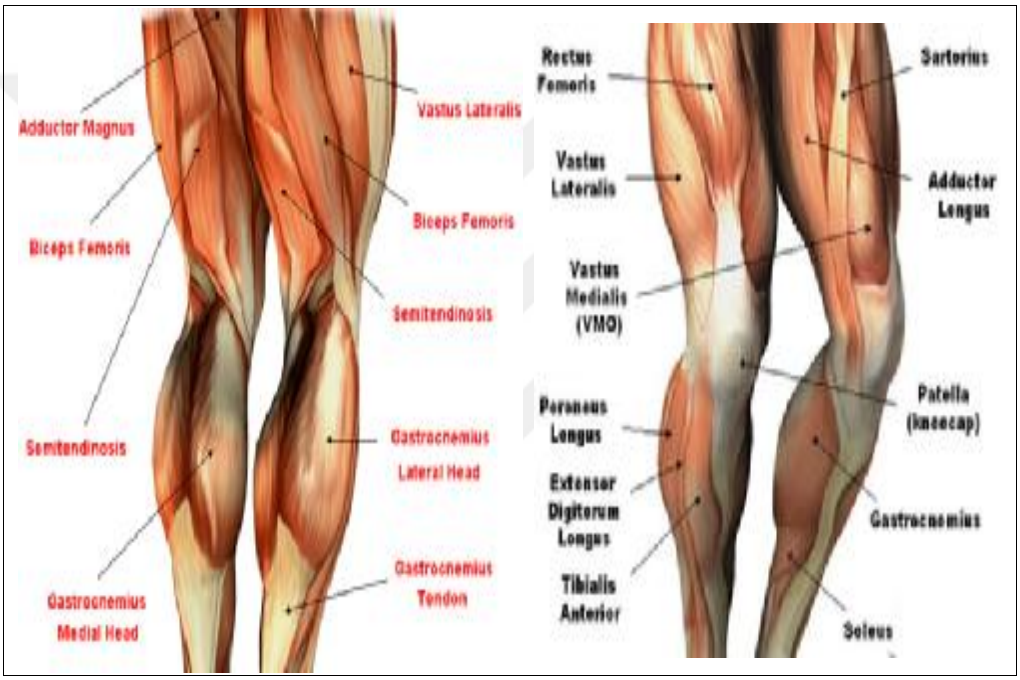


Figure 2.4. Muscles of the Lower Limb: The rear and back views (Crompton et al., 2017)

2.4. Biomedical Signal Processing Techniques

Over the past 60 years, the discipline of signal processing has been developed by the genesis of Fast Fourier Transform (FFT) and the implementation of dicrete filters, which statistically manipulate array of image, audio, and multimedia signals (Semmlow, 2004).

For most of biomedical signal processing, the information is collected in the form of image or signal of electrical change (Yi et al., 2016), (Dorf, 2006), (Elgendi, 2016). Depending upon the body organs, application or signal characteristics, the three mostly biomedical signals used for research are ECG, EMG and EEG (Sharma et al., 2016), (Akben et al., 2016) (Singh, C. and Singh, J, 2016). The real nature of these three signals is analog, but in order to process these signals using a computer, we need to convert them to the digital signals, so that the microprocessor will be able to store, read, understand and manipulate the pertaining data.

In fact, a signal is termed biological if it is captured from a living organisms, and carries instructions which tell about the attitude or character of that organism. For instance, the temperature reading of a a sick person, the voltage read by an sensor attached onto the scalp,.. are biological signals (Delgutte, 2007). Most common signal processing architectures are constisted of four parts as seen Figure 2.5.

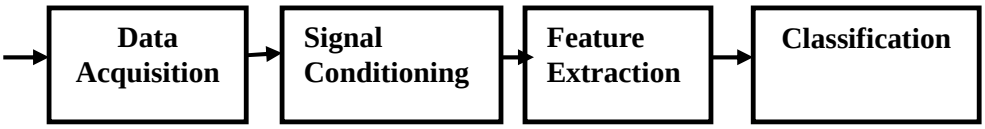


Figure 2.5. Stages of signal processing for biomedical application

In Figure 2.5, the goal of data acquisition is to take up the signal and translate it in a form that can be understood by computers for the purpose of signal manipulation or processing (Lanata et al., 2015), (Tsividis, 2010). At this phase, the main thing is to prevent from any loss of information about the signal, while that of signal conditioning like filtering is to exclude or attenuate incidental components like noises from the signal (Helleputte et al., 2015), (Clancy et al., 2002). Third stage which is Feature extraction, means identifying and measuring a small number of features (Nixon and Aguado), (Patidar et al., 2013), (He et al.,

2015). The identified features must contain enough information of signals and must be simple enough for fast training and classification. The final phase of signal processing is classification is very important in medical applications where a decision has to be made (Trinka et al., 2015), (Subasi et al., 2013). This stage intends to respond questions like whether the patient have a brain cancer or not based on a brain images or if the patient have some particular abnormality in his heart beats following the results of an electrocardiogram

3. MATERIALS AND METHODS

In this section, we describe the materials and the methods of this thesis, and they are summarized in the block diagram of the Figure 3.1.

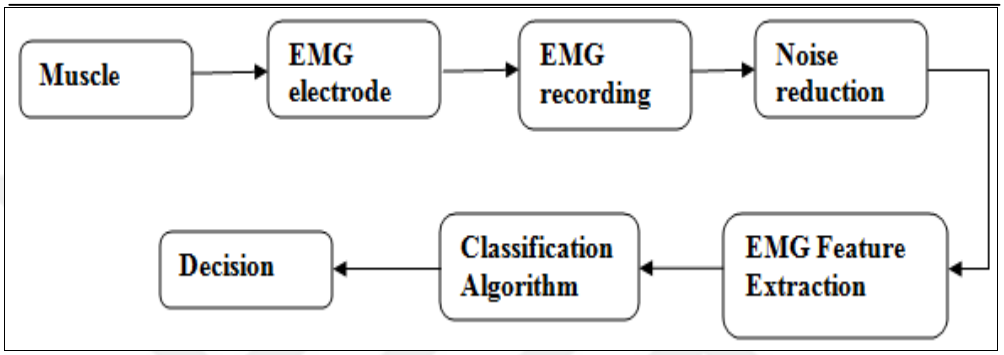


Figure 3.1. Stages of an electromyographic pattern classification system

3.1. Discription of Lower Limb EMG Datasets.

In this thesis, the datasets to be used are newly uploaded ones available at a well-known machine learning repository (Sanchez, et al., 2014). Twenty two subjects were recruited to record the datasets. Half of the subjects had diverse knee pathologies that had been previously confirmed by a doctor. Three exercices which are marching, sitting, and standing were underwent by the participants as the state relating to the knee muscles was investigated. Surface EMG electrodes were placed on four muscles crossing the knee. The muscles in question were the vastus medialis, semitendinosus, biceps femoris and rectus femoris. The equipment used to record the data was MWX8. This instrument had eight digital channels and four analog channels. The first four channels were used to record surface electromyogram and another one channel was used to read in degrees of flexion in the knee. All these data were recorded on a memory card and the resolution used was of 14-bits and the sampling frequency of 1000Hz.

3.2. Data pre-processing

According to Banhatti and Deka (2016), artificial neural networks grasp knowledge quickly and provide better accuracy if the training samples are pre-processed before being used to teach or train the model. As shown in Figure 3.2, a data pre-processor is used to change rough data to easily computable data. The aims of data pre-processing comprise size reduction of the examples space, data normalization, reduction of data impurities, feature extraction,...

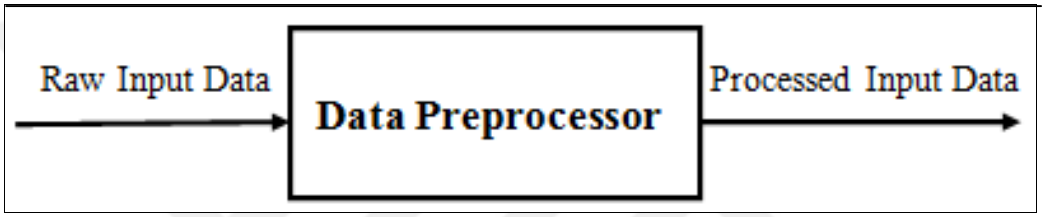


Figure 3.2. Block diagram of data preprocessing

Among these data preprocessors, Normalization, noise reduction, and feature extraction were used in this thesis.

3.2.1. Normalization

This process refers to standardizing the data dimensions so that they have almost the same range, while noise reduction refers to incresing the signal to noise ratio by using smoothing techniques (García, et al., 2016), (Wilson, 2017). There are two common ways of achieving normalization: One is standard score or Zscore normalization**.** In Zscore normalization, shown in Figure 3.3, the concept is to divide each input variable data by its standard deviation, once it has been zero-centered. The basis of Zscore normalization is on the formula below:

$$y_{new} = \frac{y_{old} - Mean}{std} \tag{3.1}$$

Where, y_{old} is the initial value, y_{new} is the final value, and $Mean$ and std are the mean and standard deviation of the initial samples, respectively.

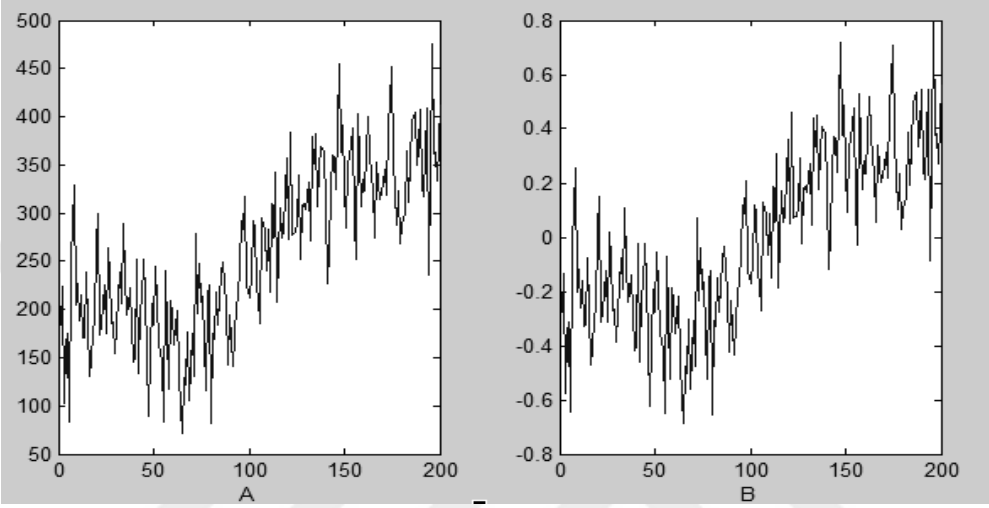


Figure 3.3. Originally unnormalized data and normalized data using Zscore normalization

Sigmoidal normalization is an another form of this pre-processing (Shedeed and Issa, 2016.). It normalizes each data sample using a sigmoid function so that the min and max along the dimension is -1 and 1 respectively. The mean and standard deviation of the training samples must be computed before anything else. The linear or semi-linear areas of this function in question relates to those examples which are found in the standard deviation of the mean, while those deviating samples called outliers are squeezed on the rear ends of the sigmoid function. The Formulas are given below:

$$y_{new} = \frac{1 - e^{-a}}{1 + e^{-a}} \tag{3.2}$$

Where,

$$a = \frac{y_{old} - Mean}{std} \tag{3.3}$$

The sigmoidal normalization is an appropriate approach used to introduce the outliers of the samples which normally present higher values. The samples plotted Figure 3.4 have two high deviating samples. The normalization with a sigmoidal function is always preferred because it captures the high outlier values while mapping the training samples to the sigmoidal scale.

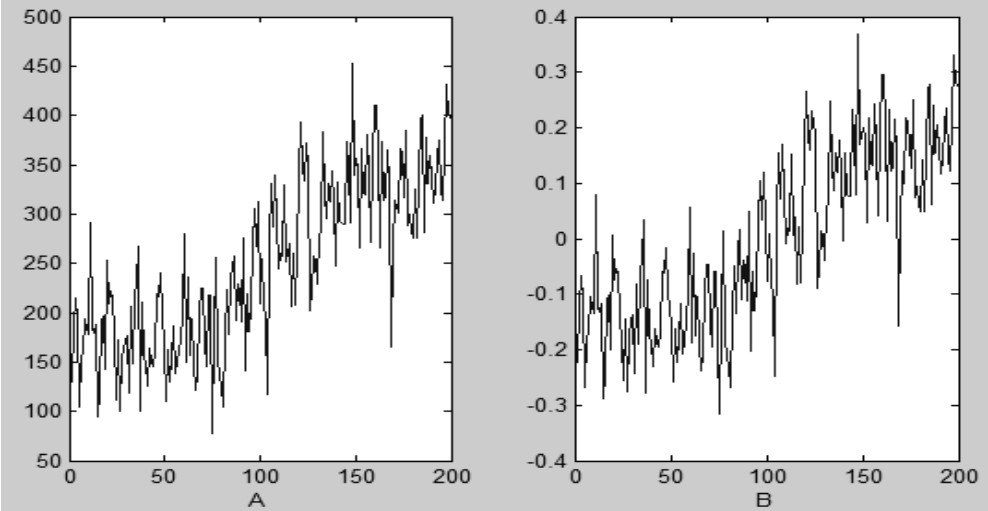


Figure 3.4. Originally unnormalized data and normalized data using sigmoidal normalization

3.2.2. EMG Signal Processing and Feature Extraction.

In order to capture the most important knowledge or information in the electromyogram, feature extraction procedure should be performed. The reaped features of the electromyogram are usually computed depending on linear or statistical methods (Phinyomark, et al., 2011). The purpose of feature extraction is to obtain the important characteristics of a data and represent them into a set of

features (Krishnan, et al., 2009). As seen in many literatures, a number of significantly extracted features depend on time-domain statistics such as the mean of the samples, the stand deviation of the samples, the variance of the sample signals, the zero crossing value, slope sign changes, which need less computational resources. The time-domain methods have abundantly been employed in many previous researches for feature extraction, due to the reason that they comparatively present no difficulties in their implementation and present high performance (Oskoei and Hu, 2008) and (Tkach, et al., 2010).

3.3. Classification Algorithms

In biomedical applications, a classification problem is considered as decision making and was previously performed by a specialised doctor. After the introduction of classification algorithms, computers could do this role all alone in an automatic way or could help doctors to make easy and accurate decisions.

In this thesis, based on extracted feature from raw SEMG signals, a number of classification algorithms were investigated and compared using Matlab codes and Weka toolbox (Hall, et al., 2009). The classification algorithms hereby employed are described in this subsection.

3.3.1. Multilayer Perceptron (MLP)

Multilayer Perceptrons (MLPs) were developed to solve XOR or non-linear problems that Rosenblatt's Perceptron could not solve (Rowcliffe et al., 2006). Figure 3.5 depicts the block diagram of MLP that has two hidden layers and one output layer. The training signals move through the structure in a progressive direction, from input layer towards an output layer.

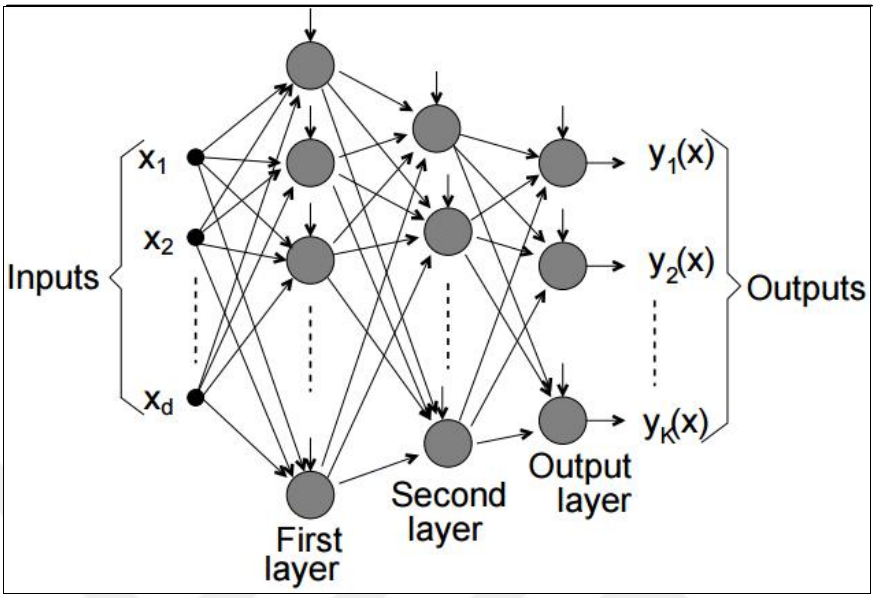


Figure 3.5. Block diagram of a multilayer perceptron with two hidden layers

The training of this network is done by the widely well-known procedures recognized as the error back-propagation algorithm (Wani, 2013). As written in the work of (Cilimkovic, 2015), the aim of back- propagation algorithm is to show the logic behind this algorithm. The concept of back- propagation algorithm is very simple, where the output of a neural network is estimated based on the needed output. If the solution is not good enough, the value of the links between the layers are adjusted and the computations are done over and over again and until the difference between the desired output and the estimated value is small enough. The five steps which constitute the perceptron learning rule or back-propagation are described here below:

First of all put,

- $x(n) = (m+1)' 1$, is the variable vector
  $= [+1, \chi_1(n), \chi_2(n),..., \chi_m(n)]^T$
- $w(n) = (m+1)' 1$, is the weight vector
  $= [b(n),, W_1(n), W_2(n),..., W_m(n)]^T$
- $b(n)$ is the bias,
- $y(n)$ is the system response,
- $d(n)$ is the needed response,
- η is the learning rate parameter, which is a constant no greater than a unit.

Where *n* is the time step, $n = 0,1,2,...$

Step 1: ***Initialization***. The weight is first set as $w(0) = 0$, *then computed for* $n = 0,1,2,...$

Step 2: ***Activation.*** *At time step* n *, the system is activated by feeding in the variable vector* $x(n)$ *and the expected response* $d(n)$.

Step 3: ***Computation of actual response.*** *The system response of the Perceptron is estimated by*

$$y(n) = signum[w^T(n)x(n)]$$

Where, signum is the activation function given by the equation 13.

Step 4: ***Adaptation of weight vector.*** *The weight vector of the network is updated using the following learning routine:*

$$w(n+1) = w(n) + \eta[d(n) - y(n)]x(n)$$

Where,

d (n) =1 if x (n) belongs to class A, or
d (n)=-1 if x(n)belongs to class B

Step 5: ***Continuation.*** *Increase the time step n by one and return to step 2.*

3.3.2. Support Vector Machine (SVM)

Support Vector Machines (SVM) learning algorithms have been mainly intended for categorizing a two class problem. They try to separate samples of different classes with a maximum margin hyperplane (Yeh, 2015). The procedures of SVM was first invented to handle limited case of categorizing training samples without errors, and they are considered as a new group of learning algorithms, as efficient, competent and popular as artificial neural networks (Cortes and Vapnik, 1995).

Given categorized training samples, the the result of SVM algorithms is an optimal hyperplane which classifies previously unseen samples. Figure 3.6, shows the idea of the optimal plane. In Figure 3.6.a, There are many possible decision boundaries that would perfectly separate the two classes. But, the support vector machine algorithm will choose the line in two-dimension (or "hyperplane", more generally) that maximizes the margin around the boundary. As seen in Figure 3.6.b, any line passing too close to the data points is a bad one and will not generalize correctly.

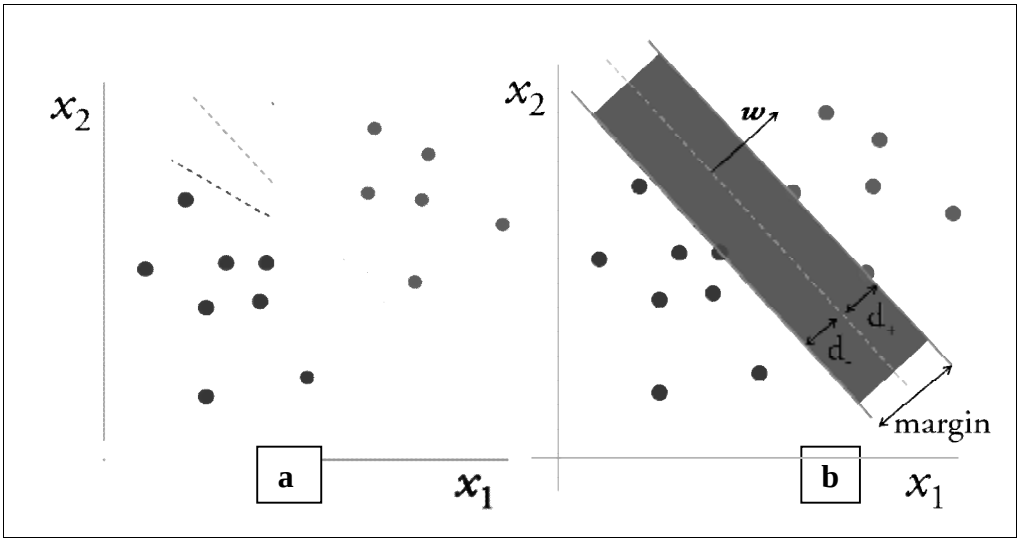


Figure 3.6. Deciding the optimal hyperplane (Altini et al., 2015)

Two types of SVMs have been considered in this thesis: Linear SVM (LSVM) and Quadratic SVM (QSVM). These algorithms have empirically showed good performance and have successful applications in many fields including bioinformatics, text, image recognition, etc. (Chamasemani and Singh, 2011).

3.3.2.1. Linear Support Vector Machines (LSVMs)

To understand the concept behind SVM algorithms, let's consider a linear classifier that has the following function:

$$f(w,b) = signum(w^T + b) \tag{3.4}$$

Where, w is the vector normal to the plane, and b is a constant that describes how much the plane is shifted relative to the origin, and , x stands for the data points which are the most adjacent the hyperplane. The distance D of the plane from the origin being

$$D = \frac{b}{\|w\|} \tag{3.5}$$

Among all the possible reprentation of the a hyperplane, let's choose

$$\left|w^T * x + b\right| = 1 \tag{3.6}$$

Now, the distance between point x and the plane is given by

$$d_+ = d_- = \frac{\left|w^T * x_i + b\right|}{\|w\|} = \frac{1}{\|w\|} \tag{3.7}$$

As from Figure 3.7, the margin is twice the distance d_+ and d_-, then the margin becomes

$$Margin = \frac{2}{\|w\|} \tag{3.8}$$

So to maximize the margin, we have to maximize $\frac{1}{\|w\|}$. However, this is an unpleasant (non-convex) objective function.

For this reason, we now have to solve the following constrained optimization, which can be solved by the Lagrangian multiplier methods because it is quadratic.

$$\min_{w,\xi_i} \frac{1}{2} w^T w + C \sum_{n=1}^{N} \max(0, 1 - y_i (w^* x_i + b)) \tag{3.9}$$

Where, in the equations (3.7)-(3.9), x_i are support vectors, w is the normal vector to the plane, y_i is the desired class response, b indicates the bias, C is the normalization constant and, ξ_i are slack variables (Koo et al., 2008). The of equation (3.9) is referred to as the primal form of norm L1-SVM.

3.3.2.2. Quadratic Support Vector Machine (QSVM)

To make a QSVM, we should make the equation (3.9) differentiable. This results in another mostly employed SVM algorithm recognized as the dual form or as the norm L2-SVM which minimizes the squared hinge loss:

$$\min_{w,\xi_i} \frac{1}{2} w^T w + C \sum_{n=1}^{N} \xi_i^2 \tag{3.10}$$

Now, the fact that norm L2-SVM is differentiable, depicts the difference in performance between QSVM and LSVM. For all kernalized support vector machines, including QSVM, optimization must be dperformed in dual form instead of primal form (Wright, 2009).

3.3.3. Logistic Regression

Logistic Regression (LR) is a way for approximating functions of the nature $f : X \to Y$, or $P(Y|X)$ in the case where Y is sequencial or digitital-valued, and $X = X_1, X_2, ...X_n$ is any input vector having discrete or continuous variables (Mitchell, 2015).

Logistic regression considers a parametric form for the distribution $P(Y|X)$ to precisely calculate its parameters from the examples. The parametric system estimated by logistic regression when Y is of digital value (0 or 1) is given by:

$$P(Y = 1|X) = \frac{1}{1 + e^{(w_o + \sum_{i=1}^{n} w_i X_i)}} \tag{3.11}$$

and

$$P(Y = 0|X) = \frac{e^{(w_o + \sum_{i=1}^{n} w_i X_i)}}{1 + e^{(w_o + \sum_{i=1}^{n} w_i X_i)}} \tag{3.12}$$

To classify any given sample X with logistic regression, we actually seek to decide the value C that maximizes $P(Y = C|X)$. Furthermore, the distribution $P(Y|X)$ is expected to trace the shape of a logistic function shown in Figure 3.7.

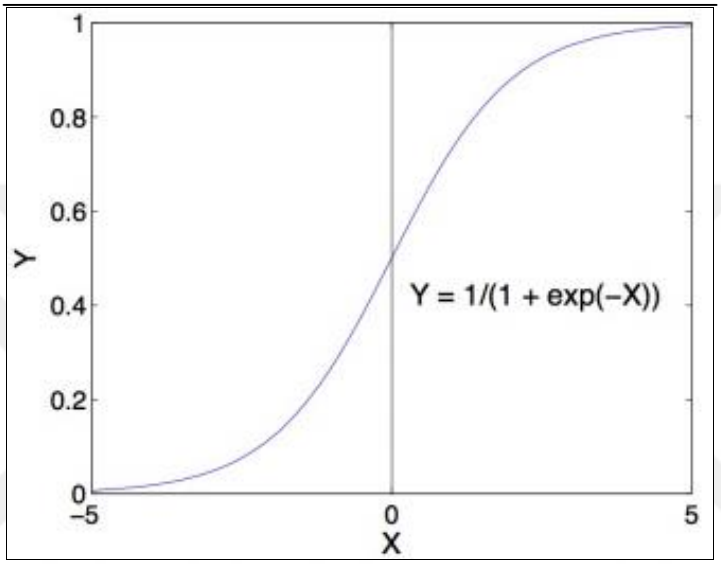


Figure 3.7. Form of the logistic function (Mitchell, 2015)

Then, from Figure 3.7, considering the natural logarithm of the two sides, it brings to a linear classification rule that attributes label $Y = 0$ when the sample X complies with the equation:

$$0 < w_o + \sum_{i=1}^{n} w_i X_i \tag{3.13}$$

and assigns $Y = 1$ otherwise.

3.3.4. Bayesian Networks

Bayesian networks (BNs) also recognized as belief networks are grouped among the probabilistic graphical models (Gal, 2008).

Given a classification task of M classes, $w_1, w_2, ..., w_M$ and an unseen sample which is described by a variable vector x, we make the M conditional probabilities $P(w_i|x)$, $i = 1,2,...,M$. These conditional probabilities are also called posteriori probabilities and they show the probability that the unseen data sample belongs to the corresponding category w_i, given that the relating variable vector assumes the value x.

So, to derive the bayes decision theory, we let w_1 and w_2 be the two categories in which the samples are found. Now let us consider that the priori probabilities $P(w_1)$ and $P(w_2)$ are given. This is a very acceptable supposition because even if the samples are unseen before, they could be computed from the available input examples. The other statistical values considered to be given are the class-conditional probability density functions $P(x|w_i)$ $i = 1,2$, , describing the distribution of the input samples in each of the classes. conditional probabilities become,

$$P(w_i|x) = \frac{P(x|w_i)P(w_i)}{p(x)} \tag{3.14}$$

where *p(x)* is the pdf of *x* and for which we have:

$$P(x) = \sum_{i=1}^{2} P(x|w_i)P(w_i) \tag{3.15}$$

The Bayes classification rule can now be stated as

If $P(w_1|x) > P(w_2|x)$, *x is classified to* w_1

If $P(w_1|x) < P(w_2|x)$ *x is classified to* w_2

3.3.5. Complex and Medium Decision Trees

Decision Trees (DTs) are decision procedures which are arranged as branches of a tree and are used to make classifiers or regression systems (Sewaiwar and Verma, 2015). The DTs are made up of three components: *Root node*, *internal node* and *leaf node*. The uppermost component or the head of the tree represents the root node or the node without a parent. The leaf node is the terminal component of the structure and it is the node without any children. Wheareas, the nodes between Root node and the leaf nodes are called the internal node or the node with at least one child. Each inner node represents test on a feature vector, each extension represents a result of the test, and each leaf node contains a class label.

Figure 3.8, shows an example of a binary DT, where each inner node has no more than two children. In the Figure 3.8, A is the root node, B, C and E are the internal nodes, and D,F,G,H,I are the leaves of the decision tree.

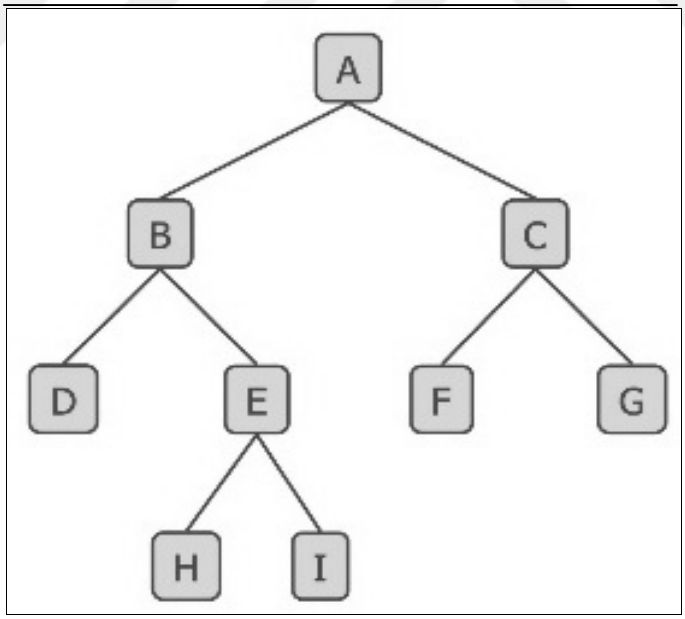


Figure 3.8. Illustrating the components of the binary decision tree

In every repetition of decision tree learning algorithms, an input sample is supplies then a representative feature is identified and serves the purpose of decomposing the whole samples into smaller data subsets; where every subset is regarded as the provided training samples for the next cycle.

In fact, th whole idea decision tree classifiers depends upon employing information gain to deduce which best feature to pick in order to test each node (Patel and Singh, 2015). To bring up the notion of information gain, we first have to know about the entropy, that estimates the quantity of real knowledge and impurities contained is a given signal. Assuming a two-categories classification problem with a set C of posive P and negative N examples, the entropy of a set C relative to this binary categorization is given by:

$$Entropy(C) = - P\log_2(P) - N\log_2(N) \tag{3.17}$$

The formula of the information gain that calculates the anticipated reduction in entropy, or uncertainty raised by dividing the examples in parts according to the selected variables, is given by

$$Gain(C, A) = Entropy(C) - \sum_{v \in Values(A)} \frac{C_v}{C} Entropy(C_v) \tag{3.18}$$

Where, $Value(A)$ is a set of all possible values for variable A and D_v is the subset of D for which variable A has a value v.

In this thesis we tested both *complex tree* and *medium tree* and the main difference between them is concluded by the amount of leaves; where for Medium Tree (MT), this amount does not go beyond twenty divisions while for Complex Tree (CT) it can go beyond 100 divisions.

3.3.6. K-Nearest Neighbors

K-Nearest Neighbors (K-NN) is a simple classification model that keeps all given training samples and use them later to categorize new samples based on a similarity measure like distance functions (Imandoust and Bolandraftar, 2013).

To categorize any unforeseen case, the k-nearest neighbors algorithm has to firstly identified the k closest neighbors from the input data to the unforeseen case. After that, a category name for the unforeseen case assigned by the K-NN model would be similar to the class label of the k closest neighboring samples. Considering that k = 1, then the unforeseen case will be automatically set to the class of its nearest neighbor.

In order to to set out k-neighbors which are closest to the new case, K-NN employs distance functions for computing the distance between the unforeseen case and a case from the training samples. The widely used distance functions to compute this distance are shown in the formulas below:

$$Euclidean = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} \tag{3.19}$$

$$Manhattan = \sum_{i=1}^{N} |x_i - y_i| \tag{3.20}$$

$$Minkowski = \left(\sum_{i=1}^{N} (|x_i - y_i|)^q\right)^{1/q} \tag{3.21}$$

Where, x is an unforeseen data point and y is a case from the the training samples.

In this thesis, by the use of the Matlab toolbox, three different K-NN algorithms: The *Fine K-NN*, *Medium K-NN* and *Coarse K-NN* were trained. All

these three algorithms employs the Euclidean learning distance to determine the nearest neighbors as expressed in the equation 3.19. The only difference between the three algorithms is that they all use different number of nearest neighbors which should be equal to 1 for Fine K-NN, 10 for Medium K-NN and 100 for Coarse K-NN.

3.3.7. Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is one the procedures widely used for both data classification and dimensionality reduction. Linear Discriminant Analysis is good in handling problems where the within-class frequencies are not the same. This method maximizes the ratio of between-class variance to the within-class variance in any specific data set thereby assuring maximal separability. To achieve this maximum separability, the inter-class distance must be large and the intra-class variances must be small

As far as LDA is sought to be employed for classification problems, the computaions are based on mahalanobis distance D_M (Mandelkow et al., 2016):

$$D_M = \sqrt{(x - m)^T C^{-1} (x - m)} \tag{3.22}$$

Where, $x = (x_1, x_2, ..., x_n)^T$ is the predictors' vector, $m = (m_1, m_2, ..., m_n)^T$ is the class centroids' vector and C is the pooled within-class covariance of the predictors.

3.3.8. Ensemble Classifiers.

Ensemble classifiers are methods for improving the performance of weak learners such as Trees (Lavanya and Rani, 2012). In fact, classification trees are adaptive and robust, but do not generalize well. The techniques of ensemble

learning used in this study improve the performance of classification trees remarkably.

3.3.8.1. Bagged Tree

Bagging, also called Bootstrap aggregating, is an ensemble learning technique designed to improve the stability and accuracy of decision tree classifiers or other learning algorithms applied in classification and regression models (Skurichina and Duin, 2002).

In this method, the base classifiers $C_1, C_2, ..., C_n$ are derived from bagging samples $X_1, X_2, ..., X_n$ respectively with replacement from the data set X. Then, final decision tree is generated as a combination of all base classifiers $C_1, C_2, ..., C_n$. Then, votes have to be collected from each classifier to a tuple in such a way to classify which class the tuple belongs to. The class label is decided for the tuple based on the maximum number of votes voted for a particular class label.

3.3.8.2. Boosted Trees

Up to nowadays, despite some new algorithms made, the most common algorithm which is used to implement boosting is adaboost (Dietterich, 2000). Contrary to Artifificial Neural Networks, The training process of adaboost does not use all the variables provided; rather, it selects only those variables known to enhance the predictive performance of the model, reducing dimensionality and potentially enhancing the training time as irrelevant features do not need to be included in training processes. Algorithm is given below (Zhu et al., 2006).

Assume teaching data set with N cases belonging to Boolean classes categorized as $f \in \{0,1\}$. The boosting algorithm is summarized below:

1. Consider initial equal weights to every cases in the original examples:

$$w_i^1 = \frac{1}{p} \quad i=1,2,...,P$$

2. For iteration $n=1,2,...,N$, set $N=10$,
   a) Randomly select a data set with P case from the initial examples using weighted resampling. The probability of picking up any given data point is dependent on its scale. *A* data point that has higher weight has more probability to be picked up.
   b) Obtain a learning classifier, $f(x)$ from the resample data set
   c) Engage the leaner $f(x)$ onto the initial example data set. If as data point is not well categorized, its $error=1$, otherwise $error=0$.
   d) Calculate the sum of the weighted errors of all input examples:

$$Error^n = \sum_{i=1}^{p} w_i^n * Error_i^n$$

   e) Compute the confidence index of the function $f(x)$:

$$\alpha^n = \frac{1}{2}\ln\left(\frac{1 - Error^n}{Error^n}\right)$$

   f) Adjust the weight values of all initial training samples using confidence index which is based on the weighted error:

$$w_i^{n+1} = w_i^n * e^{-\alpha^n * Error_i^n}$$

   If the data points are well classified, the weight values are not touched, but the weight values for misclassified data points are adjusted.
   g) After that, renormalize weight values using:

$$wi = \frac{wi}{\sum_i^p w_i^n} \quad \text{so that,} \quad \sum_i^N w_i^{n+1} = 1$$

   h) $N=n+1$, if $error<0.5$, and $n<N$, repeat steps a) through g); otherwise stop $N=n-1$
   i) After N iterations, $n=1,2,...,N$, there are N predicted classifiers $f(x)$, $n=1,2,...,N$. The final prediction for case j, is obtained by the combined prediction of the T models using voting approach:

$$y_j = sign\sum^N \alpha^n f^n(x)$$

3.3.9. Deep Neural Network.

A deep neural network (DNN) is an artificial neural network (ANN) that has a number of hidden layers of neurons between the first and the last layers (Baldi, 2012). Deep neural networks employs a series of many layers of nonlinear processing nodes for variable extraction and variable transformation (Deng and Yu, 2014), (Schmidhuber, 2015). Every data processing layer is based on the results of the previous layer and takes these results as its own input samples.

Deep networks are dominating traditional machine learning models due to their ability to handle big nonlinear data and their strong flexibility which can be modified in many ways to suit given tasks (Hinton et al., 2006), (Bengio, 2009). The pioneering results in deep learning had been available since early 2000. The DNN models at that early time employed Deep Belief Networks (DBN) to pre-train deep networks (Bengio et al., 2007). Later, similar gains achieved by use of DBN could be also obtained by Autoencoders. These autoencoders were presented for the first time in the 1980s by Rumelhart and his coworkers (1986) to address the problem of unsupervised learning.

More recently, autoencoders have regained their fame and are now being employed in the “deep architecture” approach for pre-training task (Erhan et al., 2010).

The overall process of passing knowledge to a deep learning network can be described in a series of the following steps:

- *Pre-training step*: Instruct a series of shallow autoencoders, greedily one layer at a time, using unlabeled samples.
- *Fine-tuning step A*: Instruct the final layer using labeled samples,
- *Fine-tuning step B*: Uses back-propagation to fine-tune the whole network using labled samples.

Algorithms for autoencoders are motivated by the idea of a nice representation of samples. An encoder is a deterministic mapping function that transforms an input vector *x* into hidden representation *z*. More formally, as shown in Figure 3.9, assume that the input vector to an autoencoder is represented as $x \in \mathbb{R}^{D_x}$.

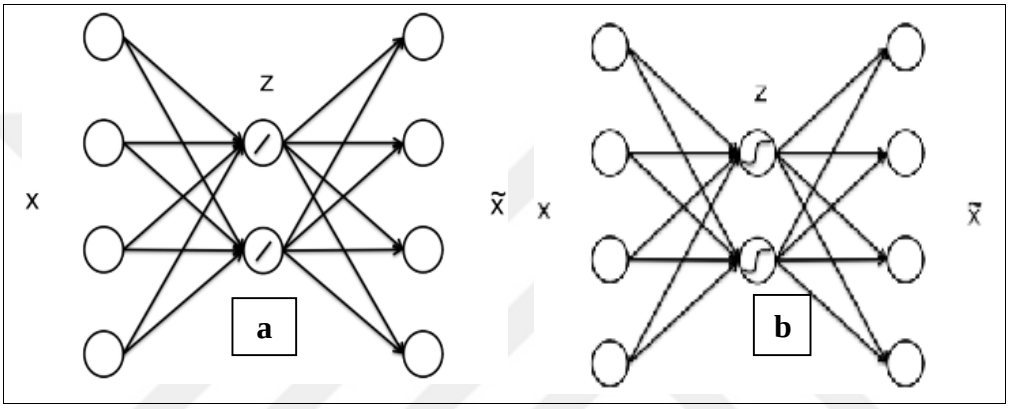


Figure 3.9. Mapping samples from 4-dimensions to 2-dimensions using a ANN with one hidden layer: a, shows a linear autoencoder and b, shows a non-linear autoencoder.

Then the encoder maps the vector *x* back and forth to another vector $z \in \mathbb{R}^{D^{(1)}}$ as follows:

$$z^{(i)} = w_1^{(i)} x + b_1^{(i)} \tag{3.23}$$

$$\hat{x}^{(i)} = w_2^{(i)} x + b_2^{(i)} \tag{3.24}$$

Where, the superscript *(i)* indicates the layer, $w_1^{(i)} \in \mathbb{R}^{D^{(i)} * D_x}$ is a weight matrix for encoder and $w_2^{(i)} \in \mathbb{R}^{D^{(i)} * D_x}$ is weight matrix for the decoder,

$b_1^{(i)}$ Î Â $^{D^{(i)}}$ is a bias vectors for encoder and $b_2^{(i)}$ and is a bias vectors for decoder. Equation 3.23 is an ***encoding algorithm*** which maps input data $x^{(i)}$ to compressed data $z^{(i)}$ and Equation 3.24 is a ***decoding algorithm*** which maps from compressed data $z^{(i)}$ back to $\hat{x}^{(i)}$ and estimates the initial data points.

So, the entire process of autoencoding is to contrast this restored input data ($\hat{x}^{(i)}$) to the initial ($x^{(i)}$) and attempts to minimize the resulting error so as to allow the restored value as similar as possible to the initial data. For this purpose, we can deduce an objective function, which is the sum of squared differences between $\hat{x}^{(i)}$ and $x^{(i)}$:

$$J(w_1, b_1, w_2, b_2) = \sum_{i=1}^{m} (\hat{x}^{(i)} - x^{(i)})^2 \tag{3.25}$$

$$= \sum_{i=1}^{m} (w_2 z^{(i)} + b_2 - x^{(i)})^2 \tag{3.26}$$

$$= \sum_{i=1}^{m} (w_2 (w_1 x^{(i)} + b_1) + b_2 - x^{(i)})^2 \tag{3.27}$$

Where, again in the formulas 3.25 to 3.27, $x^{(i)}$ are the input vectors to the autoencoder, $z^{(i)}$ are the coded input vector, $\hat{x}^{(i)}$ are the decoded input vector, w_1 and b_1 are respectively weight and bias parameters for the encoder and w_2 and b_2 are also respective weight and bias parameters for the decoder.

4. RESULTS AND DISCUSSIONS

The major purpose of this thesis was to design a medical decision support system that is intented to help doctors tell whether a visiting patient is suffering from a knee osteoarthritis or has got any other knee abnormality. The designed system could also work independently hence giving the results in an automatic manner.

To realize the system in question, electromyographic signals collected from the four muscles located around the knee joint were preprocessed and classified using machine learning algorithms. The results obtained in this thesis are given in this section which comprises four major subsections given as follows:

- Analysis results of EMG signal
- Results of filtering and noise reduction
- Results of EMG feature extraction
- Results of EMG sample classification

4.1. Analysis Results of EMG Signal

Electrical activity of EMG signal was measured and analyzed during rest, slight contraction, and forceful contraction of the muscles. To exert these actions in the muscle, the patient underwent three positions: Marching, Standing and Sitting. The waveforms associated with each muscle of interest were plotted. Both the results in normal patient and abnormal patient are shown from Figure 4.1 through Figure 4.6.

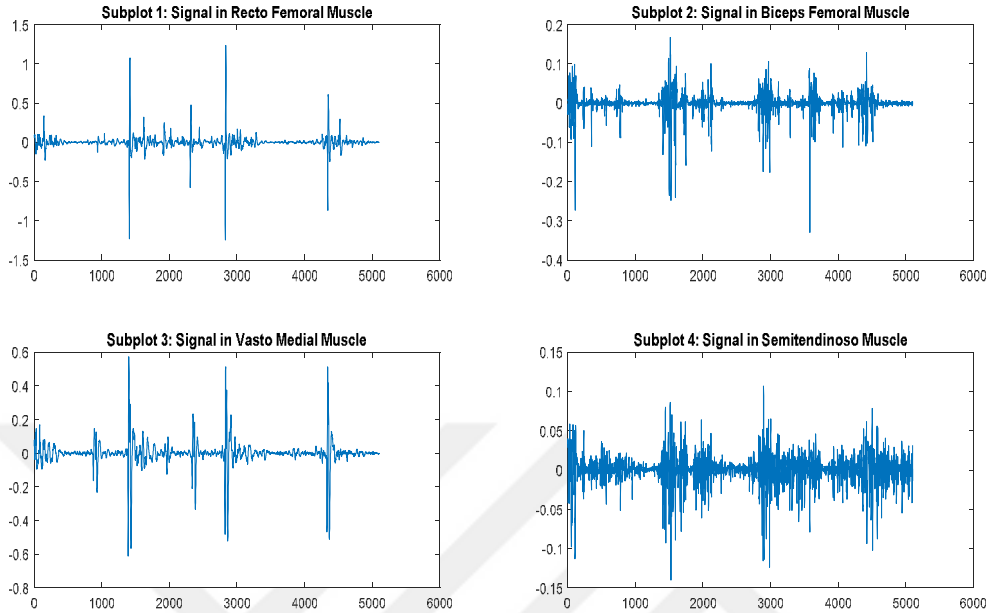


Figure 4.1. Plots of signals (Sanchez, et al., 2014) in all the four different muscles of interest for the abnormal patient in marching position

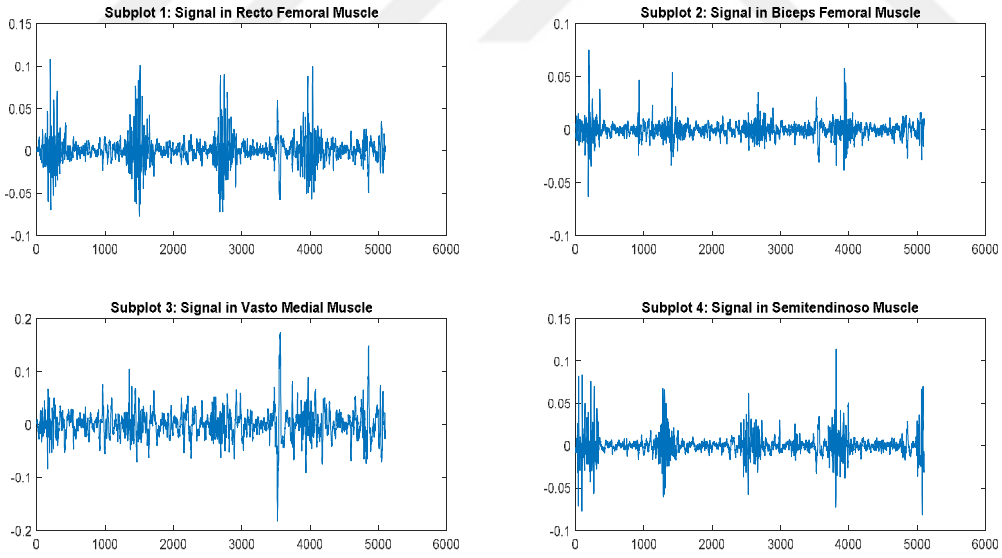


Figure 4.2. Plots of signals (Sanchez, et al., 2014) in all the four different muscles of interest for the normal patient in marching position

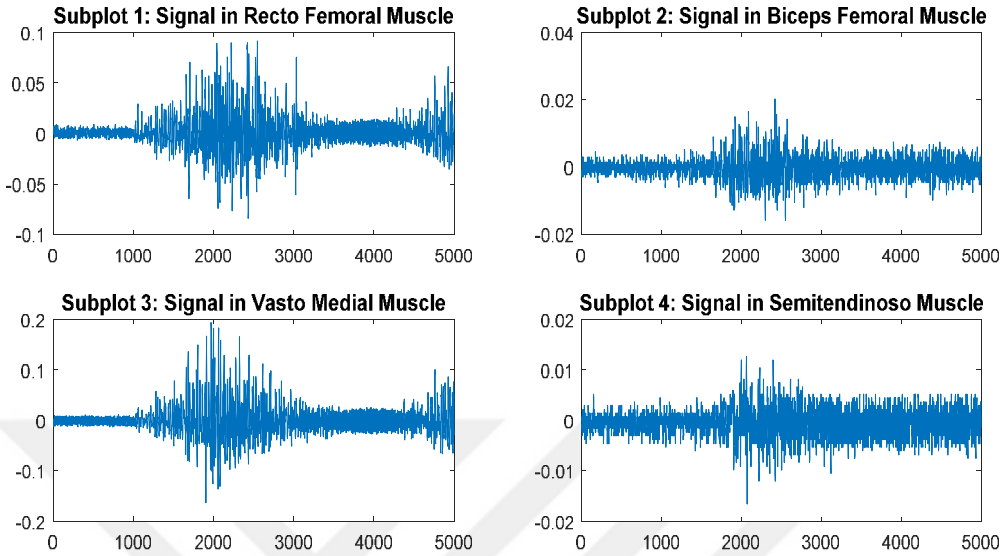


Figure 4.3. Plots of signals (Sanchez, et al., 2014) in all the four different muscles of interest for the normal patient in seated position

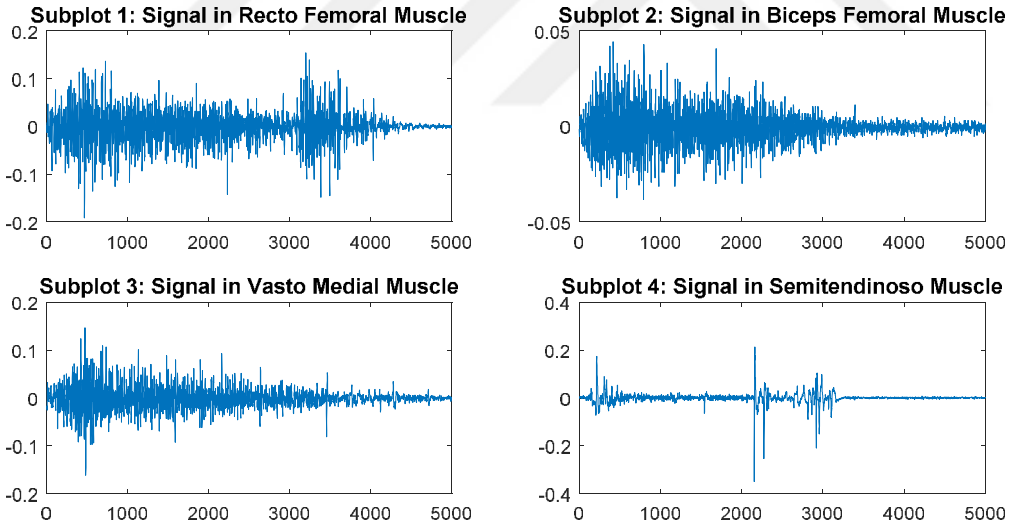


Figure 4.4. Plots of signals in all the four different muscles of interest for the abnormal patient in seated position

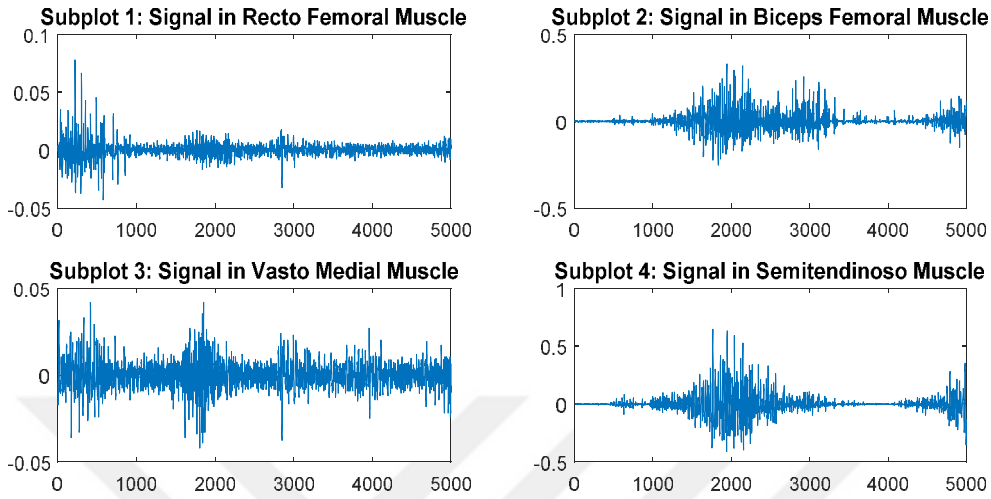


Figure 4.5. Plots of signals (Sanchez, et al., 2014) in all the four different muscles of interest for the normal patient in standing position

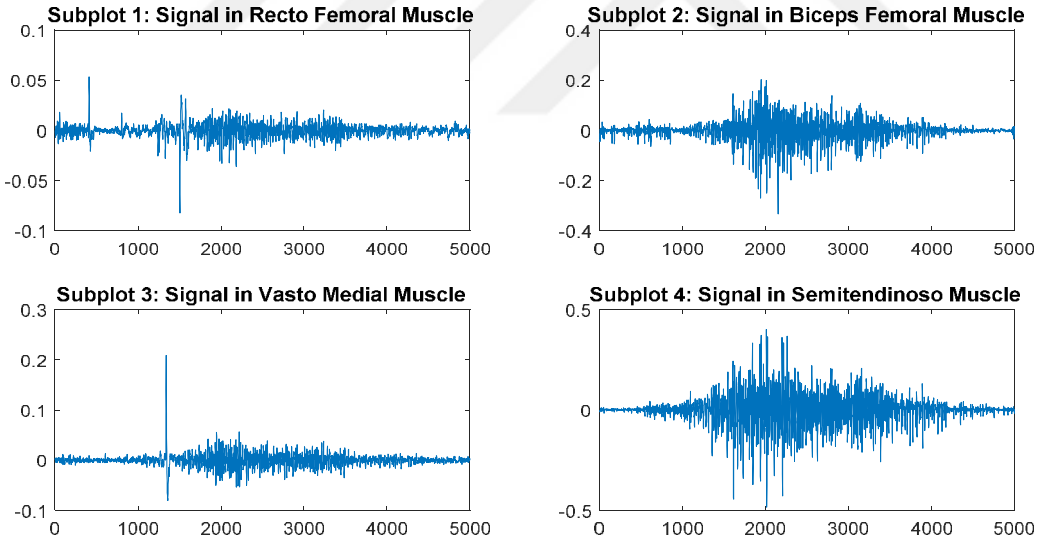


Figure 4.6. Plots of signals (Sanchez, et al., 2014) in all the four different muscles of interest for the abnormal patient in standing position

In Figure 4.1 and Figure 4.2, as normal and abnormal subjects were marching, all the four muscles exerted a certain contraction and hence, there are action potentials seen in the muscles. Actually, a healthy muscle shows almost no electrical activity during rest (Salomoni et al., 2016). However, if skeletal muscles or the bone joints are damaged, or if there is a loss of input from the nerves, a muscle may have electrical activity during rest and when it contracts, its electrical activity may produce abnormal patterns. This can be seen in Figure 4.3 through Figure 4.6, when the patient was in seated or standing postures. In these positions, it can be seen that some muscles surrounding the knee are contracted and others are at rest. For an abnormal knee, the muscles at rest: semitendinosus in a seated posture and recto femoral in a standing posture show some electrical patterns, while they almost show none when the knee has no problem.

It worth noting that these abnormal patterns are not clearly visualized in time-domain plots, however, their features may be extracted by a specific algorithm and could then be used for classification purposes.

4.2. EMG Signal Conditioning: Filtering and Noise Reduction

After having surface EMG, the signal was filtered using a 20th order Chebyshev type II band-pass filter. The bandpass filter was used to maintain the useful signal contained within the signal bandwidth by rejecting the signals whose frequencies are outside the useful frequency range (low and high frequency signals). The low frequency cutoff of bandpass filter overthrows baseline drift which is from time to time caused by slight migration of electrodes, sweating of the skin. It also suppresses any DC offset. Typical values for the low frequency cutoff range from 5 to 20 Hz (Rose, 2014). High frequency cutoff of band pass filter gets rid of high frequency noise and hinders aliasing from taking place in the sampled signal. The high frequency cutoff should be sufficiently high in order to clearly keep capturing the qwick on-off bursts of the EMG signal. Regular values for high frequency cutoff are in the range of 200Hz–1kHz (De Luca, et al., 2010).

In this thesis, the designed filter shown in Figure 4.7 and Figure 4.8 was used to remove signals that are outside the extent of the electromyogram between 20Hz and 480Hz. The filter's attenuation of unwanted signals or amplitude stop (Astop) was 40dB.

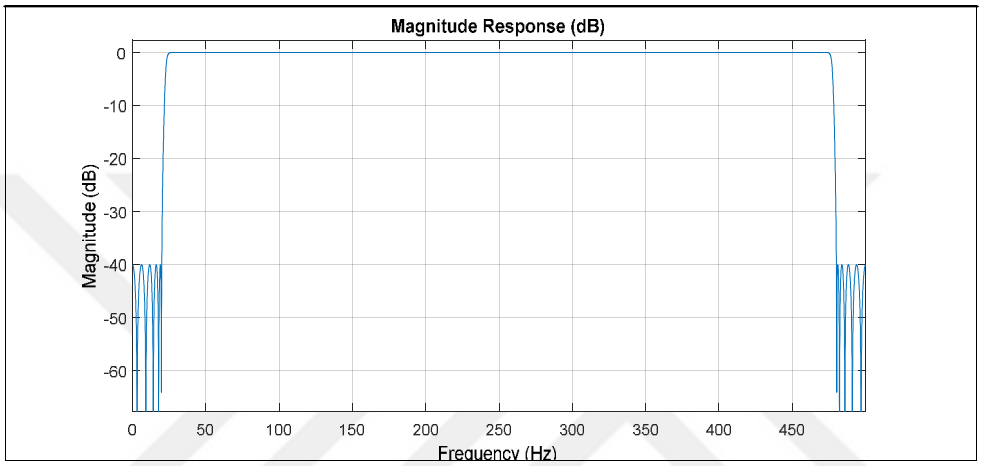


Figure 4.7. Magnitude Response of Chebyshev type II filter



Figure 4.8. Phase Response of Chebyshev type II filter

4.3. Results of EMG Feature Extraction

Feature extraction is one of the techniques used to overcome the problem of curse of dimensionality, which increases computational charges and reduces the

performance of the system due to redundant features (Kung and Mak, 2009). It is therefore important to look for means to reduce the number of features in consideration.

In this thesis, to reduce the dimensionality of SEMG data, the data were segmented using a chain of time-overlapping analysis windows as shown in Figure 4.9. The overlapping percentage of the analysis windows was 90% and the analysis window size was 5000 samples. The result of this technique was a resampled dataset with 2642 samples.

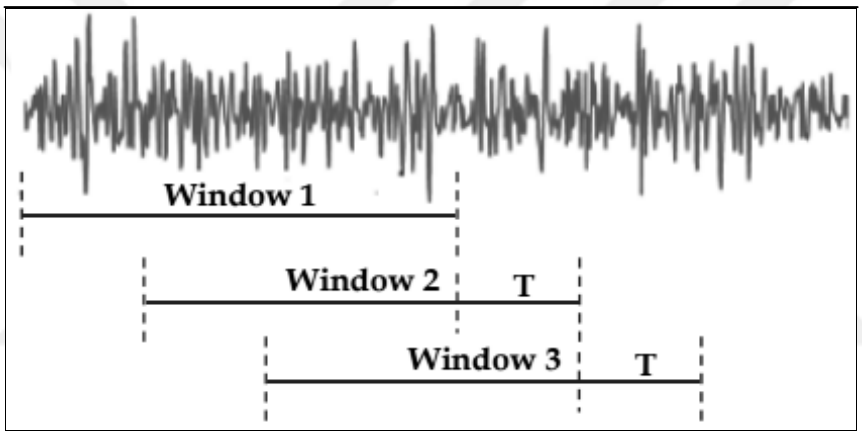


Figure 4.9. Segmentation of analysis windows of EMG recordings

Following the process of dimentionality reduction through signal segmentation technique, time domain features were extracted which resulted in a total of 30 features (6 features for each channel). The extraction of the six time-domain feactures from each channel was achieved using the below statistical formulas:

- Mean Power of the original signals:

$$P = \frac{1}{N}\sum_{n=1}^{N} x^2(n) \qquad (4.1)$$

· Mean Frequency of the original signals:

$$X(k) = \frac{1}{N}\sum_{n=1}^{N} x(n)W(n)e^{-\frac{j2\pi kn}{N}} \quad (4.2)$$

$$\bar{k} = \frac{\sum_{k=1}^{N} |X(k)|^2 * k}{\sum_{k=1}^{N} |X(k)|^2} \quad (4.3)$$

· Peak value ofthe original signal:

$$PeakValue = MAX\left(\sqrt{x^2(n) + \hat{x}^2(n)}\right) \quad (4.4)$$

Were $\hat{x}^2(n)$ is Hilbert transform of $x(n)$, where $\sqrt{x^2(n) + \hat{x}^2(n)}$ is the envelope of $x(n)$.

· Mean of the original signal:

$$m_x = \frac{1}{N}\sum_{n=1}^{N} x(n) \quad (4.5)$$

· Standard Deviation of original signals:

$$s_x = \sqrt{\frac{1}{N}\sum_{n=1}^{N} \left(x(n) - m_x\right)^2} \quad (4.6)$$

- Variance of the original signal:

$$s_x^2 = \left\{ \frac{1}{N} \sum_{n=1}^{N} \left(x(n) - m_x \right)^2 \right\} \quad (4.7)$$

4.4. Results of classification algorithms

By the use of the thirty time-domain extracted features, thirteen classifiers and seven deep learning networks were trained, tested and validated.

The performance results achieved by the first thirteen learning classifiers are described in Table 4.1, where different classifier evaluation metrics such as *Accuracy, precision, recall, F-measure* and the *error rate* were computed based on the results of confusion matrices. The formulas used to estimate these learning performance metrics are described here below:

Accuracy: Is the number of correct predictions from all the predictions made.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.8)$$

Recall: Is the number of True Positives divided by the number of True Positives and the number of False Negatives. In other words it is the number of positive predictions divided by the number of positive class values in the test data. The proportion of actual positive which are predicted positive.

$$Recall\,(Sensitivity) = \frac{TP}{TP + FN} \quad (4.9)$$

Precision: Is the number of True Positives divided by the number of True Positives and False Positives. Put another way, it is defined as the proportion of predicted positive which are actual positive.

$$Precision = \frac{TP}{TP + FP} \tag{4.10}$$

F-measure: It is also called F score and conveys the balance between the precision and the recall.

$$F\text{ - }measure = 2 * \frac{\text{Pr}ecision * \text{Re}call}{\text{Pr}ecision + \text{Re}call} \tag{4.11}$$

Error Rate: The number of mistakes made by the classifier.

$$Error\ Rate = \frac{TP}{TP + FP} \tag{4.12}$$

In the formulas, TP is True positive, TN is True Negative, FP is False Positive, and FN is False Negative.

Table 4.1. Performance results from the 13 tested learning classifiers

Classifier Group	Classifier	Accuracy%	AUC	Recall	Precision	F-measure	Error Rate
Decision Trees	*Complex Tree (CT)*	98.3	0.99	0.98	0.99	0.98	0.017
	Medium Tree (MT)	97	0.98	0.93	0.99	0.96	0.03
Discriminant Analysis	*Linear Discriminant Analysis (LDA)*	65.9	0.72	0.73	0.59	0.65	0.341
Logistic regression	*Logistic regression (LR)*	78.6	0.88	0.71	0.86	0.78	0.214
Feed Forward Networks	*MultiLayer Percetron (MLP)*	99.5	1	0.99	0.99	0.99	0.005
Bayesian Networks	*BayesNet*	98.6	1	0.97	0.98	0.97	0.014
Support Vector Machines	*Linear SVM (LSVM)*	81.2	0.85	0.71	0.93	0.8	0.188
	Quadratic SVM (QSVM)	99.4	1	0.99	0.99	0.99	0.006
Ensemble Classifiers	*Bagged Trees*	99.8	1	1	0.99	0.99	0.002
	Boosted Tree	99.4	1	0.99	0.99	0.99	0.006
Nearest Neighbor Classifiers	*Fine K-NN*	99.7	1	0.99	0.99	0.99	0.002
	Medium K-NN	98.4	1	0.98	0.98	0.98	0.016
	Coarse K-NN	83.2	0.89	0.86	0.80	0.83	0.168

Apart from performance metrics described earlier such, accuracy, precision, recall, F-measure and the error rate, a Receiver Operating Characteristic (ROC) curve is another performance measure used of interest which is also valued for select the best classifier (Matthias, 2012), (Rotello and Chen, 2016), (Majnik and Bosnic´, 2013).

In this thesis, to make comprison among the different learning models, ROC measures were translated to unique scalar value obtained by computing for a value known as *"Area Under the ROC Curve (AUC)"*. Putting in consideration all the results from these performance measures, it follows that the best three classifiers were under the group of Support Vector Machines (SVM) , Feed Forward Networks (FNN), k-nearest neighbor classifiers and ensemble classifiers. Some of the algorithms in these groups did not perform well. The most four interesting learning models for this application with their relative accuracies are Bagged Trees (99.8%), Fine K-NN (99.7%), MPL (99.5%), and Quadratic SVM (99.4%). Now, a big question is to set out the best algorithm among MLP, Ensemble, K-NN and SVM classifiers as they all have lowest error rates.

From several findings, Support Vector Machines are usually superior to MultiLayer Perceptron (Barabino et al., 1999). Firstly, because they converge on global minima while MLPs converge on local minima, meaning that MLPs essentially miss a big picture in data. Secondly, MLPs may overfit if training goes on for too long, while SVMs don't suffer from this problem.

On the other hand, MLPs have also some advantages over SVM that might make them more applicable for certain circumstances. In terms of the training time, MLP is faster than SVM and the reason for the relative slowness of SVM is that its training requires solving the associated Lagrangian dual problem rather than primal problem (Wright, 2009), (Trafalis,1999).

In contrast to both SVMs and MLPs, the main disadvantage of K-NN algorithm is that it is a lazy classifier, which means that it does not learn anything from the training data. To classify any new data sample, the K-NN

algorithm will have to first find the k closest neighbors from the training examples to the new sample. Then, the predicted class label will be set as the most common label among the k closest neighboring points. This approach leads K-NN algorithms to be slow especially when there are a large number of training instances. This can also result in a poor generalization and sensitivity to noisy data. Furthermore, K-NN does not handle soft boundaries, which is well addressed by SVM.

The Figure 4.10 and Figure 4.11 below, show the difference in terms of performance metrics among the groups of learning classifiers. It is clear that MLP outperformed with biggest measures of accuracy, recall, precision, and F-measure, yet with lowest error rate record.

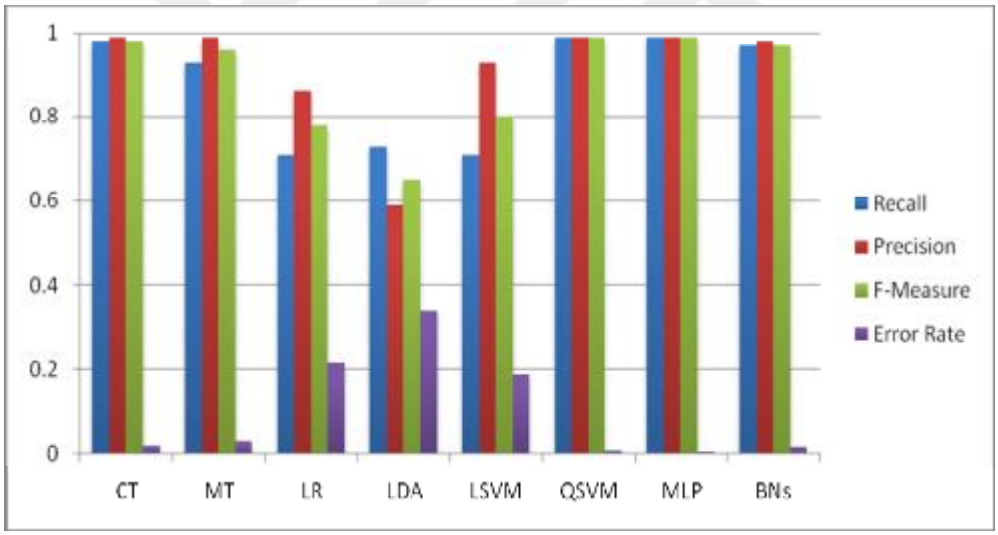


Figure 4.10. Comparing performance metrics among different groups of learning classifiers.

In addition, from Table 4.1, we can see other far more competitive algorithms from ensemble and k-Nearest Neighbor classifier groups. These are compared in column chart of Figure 4.11.

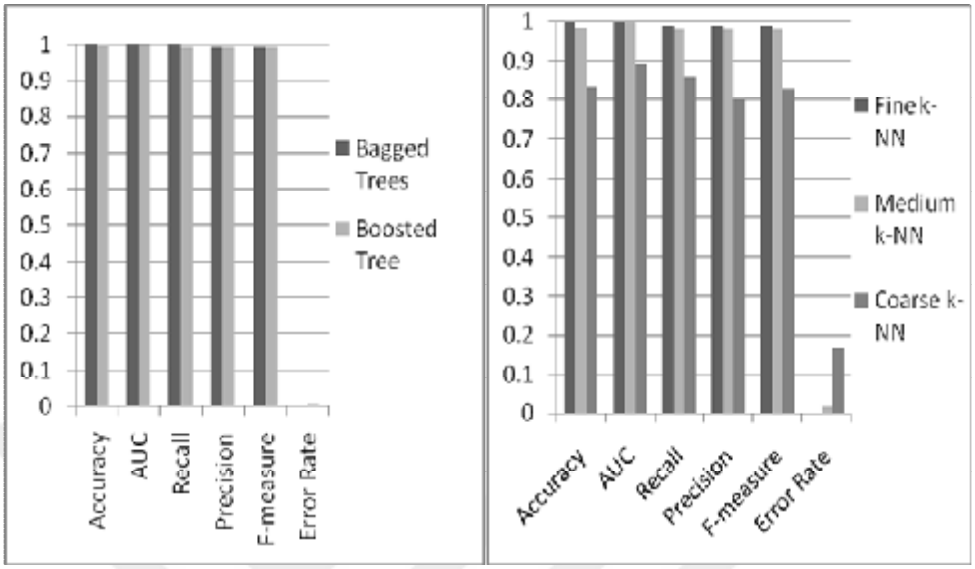


Figure 4.11. Comparing performance metrics of ensemble and K-NN classifier groups

Following the testing and validation of shallow classification models, other seven deep learning classifiers were also trained. According to previous researches and findings in the field of DNN, neural networks with multiple hidden layers have been useful for solving classification problems with complex data and have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. In this study, we tested these methods for biomedical diagnosis applications. To make a deep neural network, we firstly built an ANN model that had 2 hidden layers and trained them individually in an unsupervised fashion employing autoencoders. After that, we trained the final stage "softmax" in a supervised manner (Lecun et al., 2015). The formed deep neural network is shown in Figure 4.12.

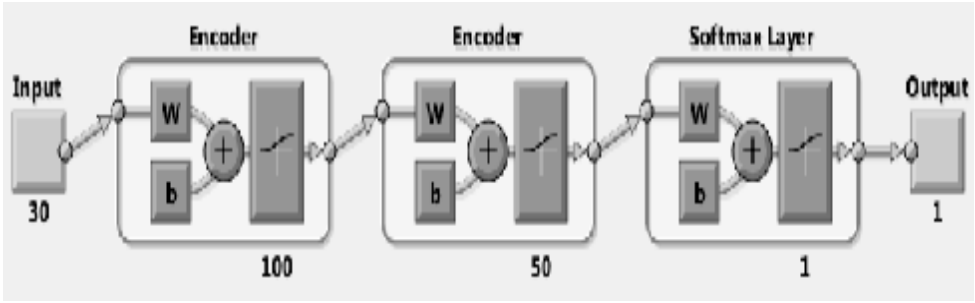


Figure 4.12. The architecture of the deep learning system

To achieve the best performance measure by the application of the implemented deep learning model, we used a wide variation of model parameters. Firstly, by adjusting the number of the hidden neurons in the sparse autoencoders, we achieved the highest accuracy of 86.6% with 900 neuronss in first layer and 200 neurons in the second layer. The results of this process is depicted in Figure 4.13 below.

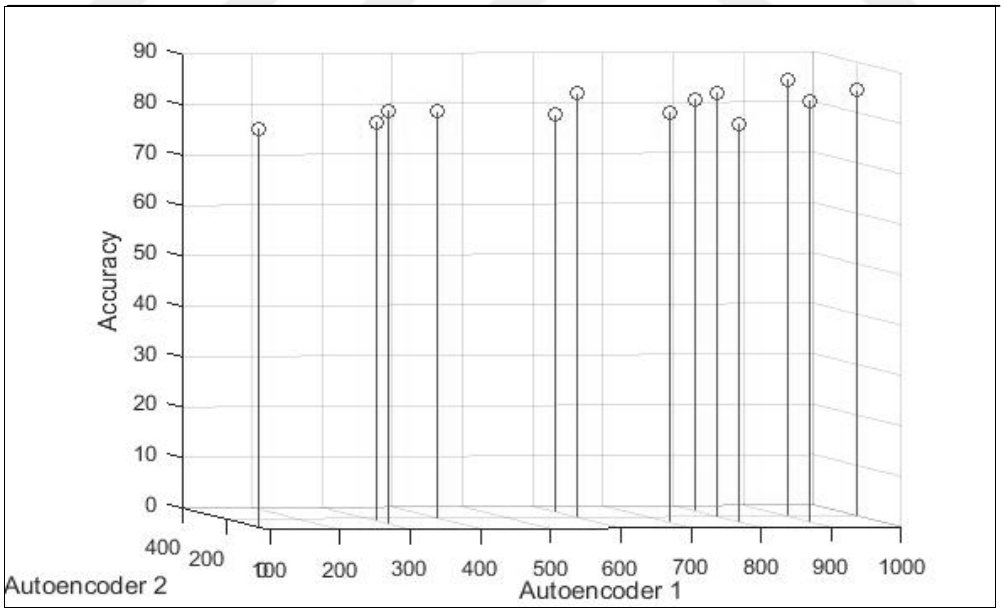


Figure 4.13. Results of the achieved accuracy by varying the number of neurons in the two autoencoders

Secondly, by adjustiong other parameters of the DNN such as *maximum epoch, sparsity regularizer* and the *sparsity proportion.* Figure 4.14, different accuracies were achieved; where the maximum performance measure about 86.6% of accuracy was reached at the maximum epoch of 1000, while the sparsity regularizer was 7 and the sparsity proportion was 0.35.

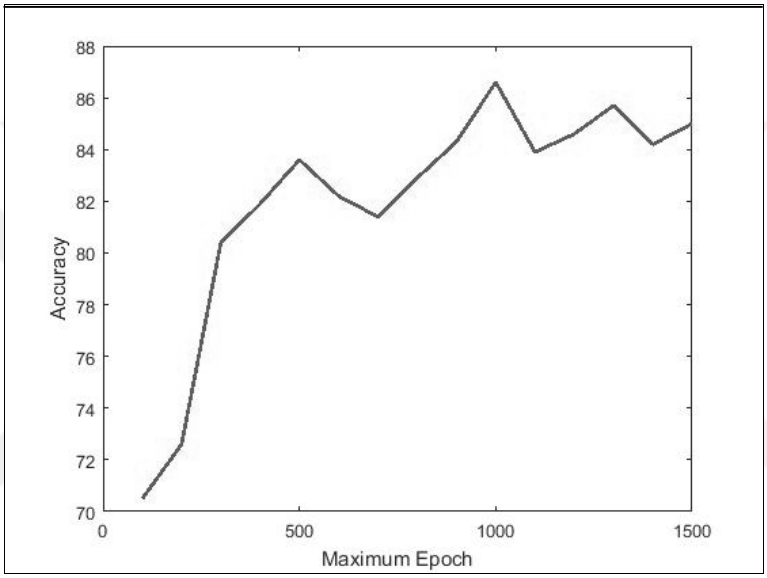


Figure 4.14. Different performance measure achieved by adjusting maximum epoch parameter of the DNN

In this thesis, there are other new methods that were attempted in order to improve the performance of the deep learning model in question. The methods were achieved by substituting the softmax algorithm with other algorithms such as *Complex Tree (CT), Medium Tree (MT), Linear Discriminant Analysis (LDA), Logistic regression (LR*) and Fine k- Nearest Neigbour (K-NN). In other words, it means that in place of supplying the results of the second sparse autoencoder to the final softmax layer, they were rather supplied to those algorithms. The highest performance result obtained with this method was an accuracy of 91.3% achieved by K-NN algorithm in the final layer.Unfolded results are shown in Table 4.2 and

the performance comparison among the new methods of DNN has been shown in Figure 4.15.

Table 4.2. Results from new methods attempted to improve the performance of the deep learning model

Base algorithm	**Final layer algorithm**	**Accuracy%**	**AUC**	**Recall**	**Precision**	**F-Measure**	**Error Rate**
Sparse Auto-encoders	*Softmax*	86.6	0.91	0.90	0.89	0.89	0.134
	CT	86.5	0.89	0.86	0.90	0.88	0.135
	MT	81.9	0.85	0.82	0.88	0.85	0.181
	LR	88.9	0.88	0.86	0.90	0.88	0.111
	K-NN	91.3	0.91	0.88	0.93	0.90	0.087
	QSVM	89.8	0.94	0.84	0.93	0.88	0.102
	LDA	64.6	0.54	0.55	0.81	0.65	0.354

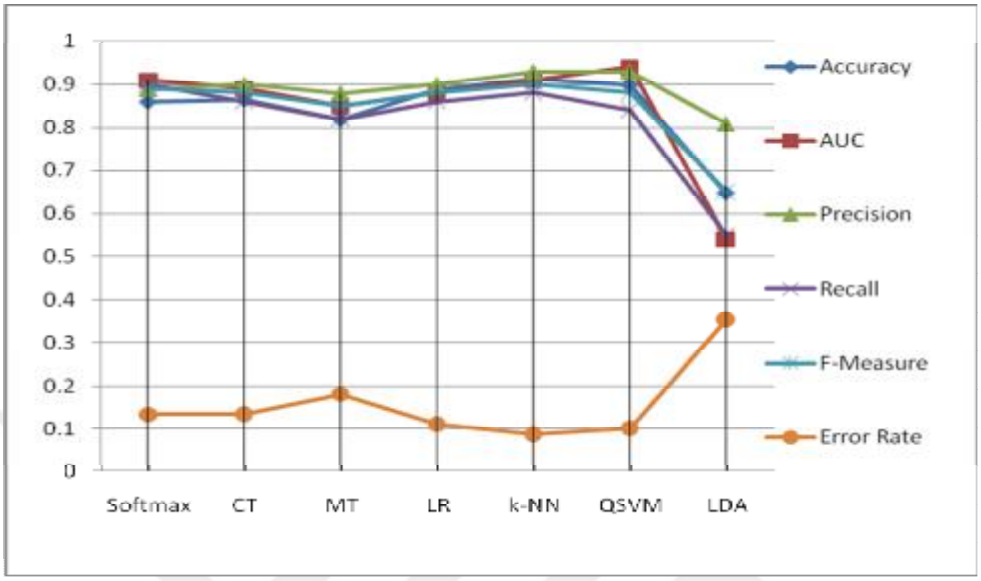


Figure 4.15. Performance comparison among the new methods of DNN

5. CONCLUSION

Automatic and computer-aided diagnosises are very important in medical field. Computer-Aided Diagnosis (CAD) systems have become very crucial for medical diagnosis of many life threats. The CAD systems enhance the diagnostic accuracy and reduce the ordinary decision-making time taken by a physician. These systems can also facilitate early diagnosis which can enhance the efficacy of treatment and increase the patient's probability of healing. Artificial Neural Network (ANN) are among the current CAD applications employed to build clinical decision support system, because they are less complex algorithms in architecture, data structure and user interface.

In this thesis, the main purpose was to investigate an ANN-based clinical decision support system that is intented to help doctors predict whether a visiting patient is suffering from a knee osteoarthritis or has got any other knee abnormality.

To realize the system in question, electromyographic signals collected from the four muscles located around the knee joint and the knee goniometric data were preprocessed and classified using machine learning algorithms.

In this study, the best accuracy above 99% was reached by the most of the investigated algorithms under the group of Support Vector Machines (SVM), MultiLayer Perceptron (MLP), k-Nearest Neighbor (K-NN) and Ensemble classifiers. In comparison, the possibility of validating the use of deep learning to solve the same problem was no better. The accuracy achieved by deep learning was 86.6% with softmax algorithm in the final layer and a performance measure of 91.3% by substituting the softmax algorithm with with k-nearest neighbour in the final layer.

In conclusion, we achieved all the set-out objectives and we showed that it is possible and of great importance to predict any knee pathology using the data

recorded by SEMG and by goniometer, then classify them using Artificial Neural Network (ANN).

REFERENCES

Ahsan, M.R., Ibrahimy, M.I., Othman O. Khalifa, O.O., 2012. The use of Artificial Neural Network in the Classification of EMG Signals. Third FTRA International Conference on Mobile, Ubiquitous, and Intelligent Computing, 225- 229.

Akben, S.B., Tuncel, D., Alkan, A., 2016. Classification of multi-channel EEG signals for migraine detection. Biomedical Research, 27, 743-748.

Alnahdi, A.H., Zeni, J.A., Mackler, L.S., 2012. Muscle Impairments in Patients With Knee Osteoarthritis. Sports Health, 4 , 284–292.

Altini, M., Casale, P., Penders, J., Velde, G.T., Plasqui, G., Amft, O., 2015. Cardiorespiratory fitness estimation using wearable sensors: Laboratory and free-living analysis of context-specific submaximal heart rates Journal of Applied Physiology, 120: 9, 1082-1096.

Amato, F., López, A., Méndez, E.M.P., Vaňhara, P., Hampl, A., Havel, J., 2013. Artificial neural networks in medical diagnosis. Journal of applied biomedicine, 11, 47–58.

Anjanı., Kubakaddi, S., 2014. Automatic Texture analysis of cartilage for early detection of osteoarthritis. International Journal of Engineering and Computer Science ISSN: 2319-7242, 3, , 6013-6016.

Aslan, U.B., Kurtça, M.P., Koçyiğit, F., Koçyiğit, A., Kuyucu, E., 2016. Effectiveness of supervised resistive exercise and home-based exercise training on lower limb muscle strength in patients with knee osteoarthritis: a long-term comparative study. Osteoarthritis and Cartilage, 24, S477–S478.

Bahr, R., Krosshaug, T., 2005. Understanding injury mechanisms: a key component of preventing injuries in sport. An international peer-reviewed journal of sport and exercise medicine, 39, 324-329.

Baldi, P., 2012. Autoencoders, Unsupervised Learning, And Deep Architectures. JMLR: Workshop and Conference Proceedings, 27, 37–50.

Banhatti, A.G., Deka, P.C., 2016. Effects of Data Pre-processing on the Prediction Accuracy of Artificial Neural Network Model in Hydrological Time Series. Urban Hydrology, Watershed Management and Socio-Economic Aspects. Volume 73 of the series Water Science and Technology Library, 265-275.

Barabino, N., Pallavicini, M., Petrolini, A., 1999.Support Vector Machines vs MultiLayer Perceptron in Particle identification. European Symposium on Artificial Neural Networks Bruges (Belgium), pp. 257-262.

Barber, S., 2007. Neural Network for beginners (Part 1 of 3). General Programming- Artificial Intelligence and Machine Learning- Neural Networks.Available at https://www.codeproject.com/Articles/16419/ AI-Neural-Network-for-beginners-Part-of

Bedson, J., Jordan, K., Croft, P., 2003. How do GPs use x-rays to manage chronic knee pain in the elderly? A case study. Ann Rheum Dis, 62, 450–454.

Bengio, Y., 2009. Learning Deep Architectures for AI. Foundations and Trends in Machine Learning, 2, 1-127.

Bennell, K.L., Hunt, M.A., Wrigley, T.V., Lim, B.W., Hinman, R.S., 2008. Role of muscle in the genesis and management of knee osteoarthritis. Rheum Dis Clin North Am, 34, 731-54.

Birdi, Y, Aurora, T., Arora, P., 2013. Study of Artificial Neural Networks and Neural Implants. International Journal on Recent and Innovation Trends in Computing and Communication, 1, 258 – 262.

Blagojevic, M., Jinks, C., Jeffery, A., Jordan, K.P., 2010. Risk factors for onset of osteoarthritis of the knee in older adults: a systematic review and meta-analysis. Osteoarthritis and Cartilage, 18, 24-33.

Caesarendra, W., Lekson, S.U., Mustaqim, K.A., Winoto, A.R., and A. Widyotriatmo, A., 2016. A classification method of hand EMG signals based on principal component analysis and artificial neural network. International Conference on Instrumentation, Control and Automation (ICA), Bandung, 22-27.

Chamasemani, F.F., Singh, Y.P., 2011. Multi-class Support Vector Machine (SVM) Classifiers- An Application in Hypothyroid Detection and Classification. Bio-Inspired Computing: Theories and Applications (BIC-TA).

Chen, H., Fuller, S.S., Friedman, C., Hersh, W., 2005. Knowledge Management, Data Mining, And Text Mining In Medical Informatics. Chapter 1, Medical Informatics. ISBN 978-0-387-25739-6.

Chen, J., Zhao, P., Liang, H., Mei, T., 2014. Motion planning for Autonomous Vehicle Based on Radial Basis Function Neural Network in Unstructured Environment. Sensors, 14, 17548-17566.

Chen, K.H., Chen, P.C., Liu, K.C., Chan, C.T., 2015. Wearable Sensor-Based Rehabilitation Exercise Assessment for Knee Osteoarthritis. Sensors, 15, 4193-4211.

Chowdhury, R.H., Reaz, M.B.I., Mohd Ali, M.A.B., Bakar, A.A.A., Chellappan, K., Chang, T.G., 2013. Surface Electromyography Signal Processing and Classification Techniques. Sensors, 13, 12431-12466.

Cilimkovic, M., 2015. Neural Networks and Back Propagation Algorithm. Institute of Technology Blanchardstown Blanchardstown Road North, Dublin 15, Ireland.

Clancy, E. A., Morın, E. L., Merletti, R., 2002. Sampling, noise-reduction and amplitude estimation issues in surface electromyography. Journal of Electromyography and Kinesiology, 12, 1–16.

Claveria, O., Monte, E., Torra, S., 2015. Multiple-input multiple-output vs. single-input single-output neural network forecasting. Document de Treball-Working Paper. AQR Research Group-IREA, University of Barcelona, Av. Diagonal 690, 08034 Barcelona, Spain.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning, 20, 273–297.

Crompton, R.H., Cummings, S.W., Tangen, C., Wood, B., 2017. Human muscle system. Encyclopædia Britannica. Available at https://www.britannica.com/ science/human-muscle-system

De Dombal F. T., Leaper D. J., Staniland J. R, Mccann A. P., Horrocks J.C., 1972. Computer-aided Diagnosis of Acute Abdominal Pain. British Medical Journal, 2, 9-13.

De Luca, C.I, 1997. The Use of Surface Electromyography in Biomechanics. Journal of Applied Biomechanics, 13, 135-163.

De Luca, C.J., Gilmore, L.D., Kuznetsov, M., Roy, S.H., 2010. Filtering the surface EMG signal: Movement artifact and baseline noise contamination. Journal of Biomechanics, 43, 1573–1579.

Delgutte, B., 2007. Biomedical Signal and Image Processing. Course materials, MIT Open Courseware (http://ocw.mit.edu), Massachusetts Institute of Technology.

Deng, L., Yu, D., 2014. Deep Learning: Methods and Applications. Foundations and Trends in Signal Processing, 7, 1–199.

Dietterich, T.G., 2000. Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. Machine Learning, 40, 139–157.

Doi, K., 2007. Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential. Comput Med Imaging Graph, 31, 198–211.

Dorf, R.C., 2006. Sensors, Nanoscience, Biomedical Engineering, and instruments. The electrical engineering handbook. Third edition. CRC Press.

Eckstein, F., Burstein, D., Link, T.M., 2006. Quantitative MRI of cartilage and bone: degenerative changes in osteoarthritis. NMR Biomed, 19, 822-54.

Elgendi, M., 2016. Terma Framework for Biomedical Signal Analysis: An Economic-Inspired Approach. Biosensors, 6, 55.

Engle, R.L., 1992. Attempt to use computers as diagnostic aids in medical decision making: A thirty–year experience. Perspectives in Biology and Medicine, 35, 207–218.

Erhan, D., Bengio, Y., Courville, A., Manzagol, P. A., Vincent, P., 2010. Why Does Unsupervised Pre-training Help Deep Learning? Journal of Machine Learning Research, 11, 625-660.

Furber, S., Temple, S., 2007. Neural systems engineering. Journal of the Royal Society Interface, 4, 193–206.

Gal, I.B., 2008. Bayesian Networks, in Encyclopedia of Statistics in Quality and Reliability, DOI: 10.1002/9780470061572.eqr089.

García, S., Luengo, J., Herrera, F., 2016. Tutorial on practical tips of the most influential data preprocessing algorithms in data mining. Knowledge-Based Systems, 98, 1–29.

Goen, A., Tiwari, D.C., 2013. Review of Surface Electromyogram Signals: Its Analysis and Applications. International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering, 7, 1429- 1437.

Gonzales, T., Hernandez, H.J., Seamon, B.A., Pennington, D., Mcintosh, V., Blackman, M.R., Love, M.O.H., 2016. The Association of Lower Extremity Strength with Step Ascent and Descent Performance in Men with Knee Osteoarthritis.

Gonzalez, M.H., Hernandez, G. M., Sotelo, J. R., Sanchez, O. A., 2015. Knee functional state classification using surface electromyographic and goniometric signals by means artificial neural networks. Ing. Univ., 19, 51-66.

González, M.H., Hernández, G.M., Sotelo, J.R., Sánchez, O.A., 2015. Knee functional state classification using surface electromyographic and goniometric signals by means artificial neural networks, Ing. Univ., 19, 51-66.

Gregson, C.L., Hardcastle, S.A., Murphy, B, A., Faber, B., Fraser, W.D., Williams, M., Smith, G.D., Tobias, J.H., 2017. High Bone Mass is associated with bone-forming features of osteoarthritis in non-weight bearing joints independent of body mass index. Science direct- Bone, 97, 306–313.

Gunn, A.A., 1976. The diagnosis of acute abdominal pain with computer analysis. J R Coll Surg Edinb, 21, 170–172.

Hall, M., Frank, E., Holmes, G., Pfahringer,B., Reutemann,P., Witten, I.H., 2009. The WEKA Data Mining Software: An Update. SIGKDD Explorations, 11:1.

Hapudeniya, M.M., 2010. Artificial Neural Networks in Bioinformatics. Sri Lanka Journal of Bio-Medical Informatics1, 104-111.

Hargrove, L.J., Simon, A.M., Young, A.J., Lipschutz, R.D., Finucane, S.B., Smith, D.G.,Todd A. Kuıken, T.A., 2013. Robotic Leg Control with EMG Decoding in an Amputee with Nerve Transfers. The England Journal of Medecine, 369, 1237-1242.

Haykin, S., 1994. Neural Networks: A Comprehensive Foundation. 1st Prentice Hall PTR Upper Saddle River, NJ, USA.

He, W., Zı, Y., Chen, B., Wu, F., He, Z., 2015. Automatic fault feature extraction of mechanical anomaly on induction motor bearing using ensemble super-wavelet transform. Mechanical Systems and Signal Processing. 54–55, 457–480.

Hebb, D. O., 1949. Organization of behavior. Journal of. Clinical Psychology. 6, 335.

Helleputte, N.V., Konıjnenburg, M., Pettine, J., Jee, D.W., Kim, H., Morgado, A., Wegberg, R.V., Torfs, T., Mohan, R., Breeschoten, A., Groot, H.D., Hoof, C.V., Yazıcıoglu, R.F., 2015. A 345 µW Multi-Sensor Biomedical SoC With Bio-Impedance, 3-Channel ECG, Motion Artifact Reduction, and Integrated DSP. In IEEE Journal of Solid-State Circuits, 50, 230-244.

Henriksen, M., Alkjaer, T., Simonsen, E.B., Bliddal, H., 2008. Experimental muscle pain during a forward lunge-the effects on knee joint dynamics and electromyographic activity. British Journal of Sports Medicine, 43, 503-507.

Hinton, G. E., Osindero, S., Teh, Y. W., 2006. A fast learning algorithm for deep belief nets. Neural computation, 18, 1527-1554.

Hopfield, J. J., 1982. Neural networks and physical systems with emergent collective computational abilities. Proc. Nat. Acad. Set., 79, 2554-2558.

Imandoust, S. B., Bolandraftar, M., 2013. Application of k-Nearest Neighbor (K-NN) Approach for Predicting Economic Events: Theoretical Background. Int. Journal of Engineering Research and Applications, 3, 605-610.

Janidarmian, M., Radecka, K., Zılıc, Z., 2014. Automated diagnosis of knee pathology using sensory data. 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH), Athens, 95-98.

Kanneppady, S.S., Kanneppady, S.K., 2015. Investigations in Osteoarthritis. Global Journal For Research Analysis, 4, Issue-5,

KAramzadeh, K., Askarifar, H., Moharrami, H., 2013. Biomedical Engineering Based Artificial Neural Network. International Journal of Scientific & Engineering Research, 4, Issue 6.

Kean, C.O., Hinman, R.S., Wrigley, T.V., Lim, B.W., Bennell, K.L., 2017. Impact loading following quadriceps strength training in individuals with medial knee osteoarthritis and varus alignment. Clinical Biomechanics, 42, , 20–24.

Kernozek, T.W., Ragan, R.J., 2008. Estimation of anterior cruciate ligament tension from inverse dynamics data and electromyography in females during drop landing. Clinical Biomechanics, 23, 1279–1286.

Kıguchi, K., Hayashi,Y., 2012. An EMG-Based Control for an Upper-Limb Power-Assist Exoskeleton Robot. In IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 42, 1064-1071.

Koçer, S., Tümer, A.E., 2016.Classifying neuromuscular diseases using artificial neural networks with applied Autoregressive and Cepstral analysis. Neural Computing and Applications, 1–8,

Koo, J.Y., Lee, Y., Kim, Y., Park, C. (2008). Representation of the Linear Support Vector Machine. Journal of Machine Learning Research, 9, 1343-1368.

Krishnan, N. C., Juıllard, C., Colbry, D., & Panchanathan, S., 2009. Recognition of hand movements using wearable accelerometers. Journal of Ambient Intelligence and Smart Environments, 1, 143-155.

Kruger, R.P., Thompson, W.B., Turner, A.F., 1974. Computer diagnosis of pneumoconiosis. IEEE Transactions on Systems, Man, and Cybernetics, 4, 44–47.

Kruger, R.P., Towns, J.R., Hall, D.L., 1972. Automated radiographic diagnosis via feature extraction and classification of cardiac size and shape descriptors. IEEE Transactions on Biomedical Engineering, 19, 174–186.

Kumar, K., 2012. Artificial Neural Networks for Diagnosis of Kidney Stones Disease. I.J. Information Technology and Computer Science. 7, 20-25.

Kung, S.Y., Mak, M.W., 2009. Machine Learning in Bioinformatics, Chapter 1: Feature Selection for Genomic and Proteomic Data Mining, John Wiley & Sons, Hoboken, NJ, USA.

Lakra, S., Prasad, T.V., Ramakrishna, G., 2012. The Future of Neural Networks. Proceedings of the 6th National Conference: Computing For Nation Development, Bharati Vidyapeeth's Institute of Computer Applications and Management, New Delhi, February 23 – 24.

Lanata, A., Valenza, G., Nardell, M., Gentili, C., Scilingo,E. P., 2015. Complexity Index From a Personalized Wearable Monitoring System for Assessing Remission in Mental Health. In IEEE Journal of Biomedical and Health Informatics, 19, 132-139.

Lavanya, D., Ranı, K.U., 2012. Ensemble Decision Making System for Breast Cancer Data. International Journal of Computer Applications, 51, No.17.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Review on deep learning. Nature, 521, 436-444.

Lementowski, P. W., Zelıcof, S. B., 2008. Obesity and Osteoarthritis. American Journal of Orthopedics, 37, 148-151.

Lisboa, P. J.G., Vellido, A., Martin, J.D., 2010. Computational Intelligence in biomedicine: Some contributions. ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), April, 28-30.

Liu, X., Zhang, M., Rıchardson, A., Lucas, T., Spiegel, J.V.D., 2016. The Virtual Trackpad: an Electromyography-based, Wireless, Real-time, Low-Power, Embedded Hand Gesture Recognition System using an Event-driven Artificial Neural Network. In IEEE Transactions on Circuits and Systems II: Express Briefs, 99, 1-11.

Lodwick, G.S., Haun, C.L., Smith, W.E., 1963. Computer diagnosis of primary bone tumor. Radiology, 80, 273–275.

Majnik, M., Bosnic´, Z., 2013. ROC analysis of classifiers in machine learning: A survey. Intelligent Data Analysis, 17, 531–558.

Maltarollo, V. G., Honório, K.M., Da Silva, A.B.F., 2013. Applications of Artificial Neural Networks in Chemical Problems, Artificial Neural Networks - Architectures and Applications, Available from: https://www.intechopen.com/books/artificial-neural-networks-architectures-and-applications/applications-of-artificial-neural-networks-in-chemical-problems.

Mandelkow, H., Zwart, J. A. D., Duyn, J. H., 2016. Linear Discriminant Analysis Achieves High Classification Accuracy for the bold fMRI Response to Naturalistic Movie Stimuli. Front Hum Neurosci, 10, 128.

Mankar, V.R., 2011. EMG Signal Noise Removal Using Neural Networks. Advances in Applied Electromyography. ISBN 978-953-307-382-8.

Matthias, K., 2012. Performance Measures in Binary Classification. International Journal of Statistics in Medical Research, 1, 79-81.

Mcalındon, T.E., Bannuru, R.R., Sullıvan, M.C., Arden, N.K., Berenbaum, F., Zeınstra, S.M.B., Hawker, G.A., Henrotın, Y., Hunter, D.J., Kawaguchı, H., Kwoh, K., Lohmander, S., Rannou, F., Roos, E.M., Underwood, M., 2014. OARSI Guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis and Cartilage, 22, 363-388.

Mcculloch, W.S., Pitts, W.H., 1943. A Logical Calculus of the Ideas Immanent in Nervous Activity, Bulletin of Mathematical Biophysics, 5, 115-137.

Mcwilliams, D.F., Leeb, B.F., Muthuri, S.G., Doherty, M., Zhang, W., 2011. Occupational risk factors for osteoarthritis of the knee: A meta-analysis. Osteoarthritis and Cartilage, 19, 829-839.

Merletti, R., De Luca, C.J., 1989. New Techniques in surface electromyography. In: Desmedt J.E. (Ed.). Computer-Aided Electromyography and Expert Systems, Elsevier, 115-124.

Mitchell, T.M., 2015. Generative And Discriminative Classifiers: Naive Bayes and Logistic Regression, Machine Learning, Chapter 3. available at https://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg. pdf.

Murphy, L., Schwartz, T.A., Helmick, C.G., Renner, J.B., Tudor, G., Koch, G., Dragomir, A., Kalsbeek, W.D., Luta, G. And Jordan, J.M., 2008. Lifetime risk of symptomatic knee osteoarthritis. Arthritis & Rheumatism, 59, 1207–1213.

Myers, P.H., Nice, C.M., Becker, H.C., 1964. Automated computer analysis of radiographic images. Radiology, 83, 1029–1033.

Nagano, Y., Ida, H., Akai, M., Fukubayashi, T., 2011. Effects of jump and balance training on knee kinematics and electromyography of female basketball athletes during a single limb drop landing: pre-post intervention study. Sports Medicine, Arthroscopy, Rehabilitation, Therapy & Technology: SMARTT, 3, 14.

Nixon, M.S., Aguado, A.S., 2012. Feature extraction and image processing for computer vision. Third edition, Elsevier, 632p.

Oskoei, M. A., HU, H., 2008. Support vector machine based classification scheme for myoelectric control applied to upper limb. IEEE Transactions on Biomedical Engineering, 55, 1956–1965.

Park, S.K., Kobsar, D., Ferber, R., 2016. Relationship between lower limb muscle strength, self-reported pain and function, and frontal plane gait kinematics in knee osteoarthritis. Clinical Biomechanics, 38, 68–74.

Patel, N., Singh, D., 2015. An Algorithm to Construct Decision Tree for Machine Learning based on Similarity Factor. International Journal of Computer Applications, 111, No 10.

Patidar, M., Jain, N., Parikh, A., 2013. Classification of Normal and Myopathy EMG Signals using BP Neural Network. International Journal of Computer Applications (0975 – 8887) , 69, No.8.

Perwej, Y., Parwej, F., 2012. A Neuroplasticity (Brain Plasticity) Approach to Use in Artificial Neural Network . International Journal of Scientific & Engineering Research, Volume 3, Issue 6.

Peterfy, C., Kothari, M., 2006. Imaging osteoarthritis: magneticresonance imaging versus x-ray. Curr Rheumatol Rep, Volume 8, 16-21.

Phinyomark, A., Phukpattaranont, P., Limsakul, C., Phothisonothai, M., 2011. Electromyography (Emg) Signal Classification Based On Detrended Fluctuatıon Analysis. Fluctuation and Noise Letters, 10, 281–301.

Podraza, J.T., White, S.C., 2010. Effect of knee flexion angle on ground reaction forces, knee moments and muscle co-contraction during an impact-like deceleration landing: Implications for the non-contact mechanism of ACL injury. The Knee, 17, 291-295.

Pollard, C.D., Sigward, S.M., Powers, C.M., 2010. Limited hip and knee flexion during landing is associated with increased frontal plane knee motion and moments Clinical Biomechanics, 25, 142-146.

Pratap, K., Shelja, 2013. Artificial Neural Network (Ann) Inspired From Biological Nervous System. International Journal of Application or Innovation in Engineering & Management (IJAIEM) , 2, Issue 1.

Pullman, S.L., Goodin, D.S., Marquinez, A.I., Tabbal, S., Rubin, S., 2000 .Clinical utility of surface EMG: Report of the therapeutics and technology assessment subcommittee of the American Academy of Neurology. Neurology, 55, 171-177.

Quintana, J.M., Arostegui, I., Escobar, A., Azkarate, J., Goenaga, J.I., Lafuente, I., 2008. Prevalence of Knee and Hip Osteoarthritis and the Appropriateness of Joint Replacement in an Older Population. Arch Intern Med. 2008; 168, 1576-1584.

Robert, E.M., Stephen, J.M., Marsha, E.M., Angelo, J. C., Matthew, D., 2008. Management of the Patient with an ACL/MCL Injured Knee. N Am J Sports Phys Ther, 3, 204–211.

Roberto, M., Alberto, B., Amedeo, T., Enrıco, M., Alessandro, M.M., 2009. Technology and instrumentation for detection and conditioning of the surface electromyographic signal: State of the art. Clinical Biomechanics, 24, 122–134.

Rose, W., 2014. Electromyogram analysis. KAAP686 Mathematics and Signal Processing for Biomechanics. Updated 2011-04-27, 2012-04-13, 2014-02-19, 2014-07-23.

Rotello, C. M., Chen, T., 2016. ROC curve analyses of eyewitness identification decisions: An analysis of the recent debate. Cognitive Research: Principles and Implications, 2016, 1-10.

Rowclıffe, P., Feng, J., Buxton, H., 2006. Spiking Perceptrons. IEEE Transactions On Neural Networks, 17, 803-807.

Rudroff, T., 2008. Kinesiological Fine Wire EMG. A practical introduction to fine wire EMG applications. Version 1.0 December 2008, Printed by Noraxon U.S.A, Inc., 71p.

Rumelhart, D. E., Hinton, G. E., Williams, R. J., 1986. Learning Representations By Back-Propagating Errors. Nature, 323, 533–536.

Salomoni, S., Tucker, K., Hug, F., Mcphee, M., Hodges, P., 2016. Reduced Maximal Force during Acute Anterior Knee Pain Is Associated with Deficits in Voluntary Muscle Activation. PLoS ONE. 11, No. 8.

Sanchez, O. F. A., Sotelo, J. L. R., Gonzales, M. H., Hernandez, G. A. M., 2014. EMG dataset in Lower Limb. UCI Machine Learning repository, available on : 2014-02-05.

Schmidhuber, J., 2015. Deep Learning in Neural Networks: An Overview. Neural Networks. 61: 85–117.

Semmlow, J.L., 2004. Biosignal and Biomedical Image Processing MATLA B-Based Applications. Robert Wood Johnson Medical School New Brunswick, New Jersey, U.S.A.

Sewaiwar, P., Verma, K. K., 2015. Comparative Study of Various Decision Tree Classification Algorithm Using WEKA. International Journal of Emerging Research in Management &Technology, 4, 87-91.

Shamır, L., Lıng, S. M., Scott, W.W., Bos, A., Orlov, N., Macura, T. J., Eckley, D. M., Ferrucci, L., Goldberg, I.G., 2009. Knee X-ray image analysis method for automated detection of Osteoarthritis. IEEE transactions on bio-medical engineering, 56, 407-415.

Sharma, S., Mohali, L., Farooq, H., Chahal, N., 2016. Feature Extraction and Classification of Surface EMG Signals for Robotic Hand Simulation. Communications on Applied Electronics (CAE), 4, No.2.

Shedeed, H.A., Issa, M.F., 2016. Brain-EEG Signal Classification Based on Data Normalization for Controlling a Robotic Arm. International Journal of Tomography and Simulation. 29, No.1.

Sher, P.P., 1980. Mathematical and computer assisted procedures in clinical decision making. Human Pathology. 11, 420-423.

Shıri, R., Karppinen, J., Arjas, P.L., Solovieva, S., Juntura, E.V., 2009. The Association between Obesity and Low Back Pain: A Meta-Analysis. American Journal of Epidemiology, 171, 135-154.

Shortliffe, E.H., Cimino, J.J., 2014. Computer Applications in Health Care and Biomedicine. Fourth Edition. Biomedical informatics. Doi: 10.1007/978-1-4471-4474-8.

Singh, C., Singh, J., 2016. Biomedical Signal Processing, Artificial Neural Network: A Review. Indian Journal of Science and Technology, 9, 0974-5645.

Skurichina, M., Duin, R.P.W., 2002. Bagging, Boosting and the Random Subspace Method for Linear Classifiers.Pattern Analysis & Applications, 5, 121–135.

Smits, J. R. M., Melssen, W. J., Buydens, L. M. C., Kateman, G., 1994. Using artifi□ cial neural networks for solving chemical problems. Part I. Multi-layer feed-forward networks. Chemom. Intell. Lab., 22, 165-189.

Srikanth, V.K., Fryer, J.L., Zhai, G., Winzenberg, T.M., Hosmer, D., Jones, G., 2005. A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. Osteoarthritis and Cartilage, 13, 769-781.

Steindler, A., 1953. A historical review of the studies and investigations made in relation to human gait. J.Bone Joint Surg, 540-542.

Stiles, J., Jernigan T.L., 2010. The Basics of Brain Development. Neuropsychol Rev, 20, 327–348.

Subası, A., 2013. Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders. Computers in Biology and Medicine, 43, 576–586.

Sundaraj, K., 2009. Biceps Activity EMG Pattern Recognition Using Neural Networks. Proceedings of the 1st WSEAS International Conference on SENSORS and SIGNALS (SENSIG '08), 11-15.

Szolovits, P., Patil, R.S., Schwartz, W.B., 1988. Artificial Intelligence in Medical Diagnosis. ANNALS OF INTERNAL MEDICINE ,108, 80-87.

Teitel, A.D., Zieve, D., 2010. Osteoarthritis. MedlinePlud Medical Encycolpedia. National Institutes of Health. Last updated: Sept 26, 2011.

Tkach, D., Huang, H., Kuıken, T. A., 2010. Study of stability of time-domain features for electromyographic pattern recognition. Journal of NeuroEngineering and Rehabilitation, 7, No. 21.

Toıvanen, A.T., Helıövaara, M., Impıvaara, O., Arokoskı, J.P.A., Knekt, P., Lauren, H., Kröger, H., 2009. Obesity, physically demanding work and traumatic knee injury are major risk factors for knee osteoarthritis—a population-based study with a follow-up of 22 years. Rheumatology (Oxford), 49, 308-314.

Toriwaki J, Suenaga Y, Negoro T, 1973. Pattern recognition of chest x-ray images. Computer Graphics and Image Processing. 2, 252–271.

Trafalis, T.B, 1999. Primal-Dual Optımızatıon Methods In Neural Networks And Support Vector Machines Training. Available at http://www.svms.org/training /Traf99.pdf.

Tran, G., Smith, T.O., Grice, A., Kıngsbury, S.R., Mccrory, P., Conaghan, P.G., 2016. Does sports participation (including level of performance and previous injury) increase risk of osteoarthritis? A systematic review and meta-analysis. British Journal of Sports Medicine, 50, 1459–1466.

Trinka, E., Cock, H., Hesdorffer, D., Rossetti, A. O., Scheffer, I. E., Shinnar, S., Shorvon, S., Lowenstein, D. H., 2015. A definition and classification of status epilepticus – Report of the ILAE Task Force on Classification of Status Epilepticus. Epilepsia, 56: 1515–1523.

Tsivids, Y., 2010. Event-Driven Data Acquisition and Digital Signal Processing- A Tutorial. In IEEE Transactions on Circuits and Systems II: Express Briefs, 57, 577-581,

Wani, S.M.A., 2013. Comparative Study of Back Propagation Learning Algorithms for Neural Networks. International Journal of Advanced Research in Computer Science and Software Engineering Research, 3, 1151-1156.

WEBMED, 2017. Structures of the Knee.Available at http://www.webmd.com/ pain-management/knee-pain/structures-of-the-knee

Werbos, P.J., 1974. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. PhD thesis, Harvard University.

WILLIAMS, J.M., 2015. The use of surface electromyography within equine performance analysis. University of the West of England. Available from: http://eprints.uwe.ac.uk/25786.

Wilson, D.H., Wilson, P.D., Walmsley, R.G., Horrocks, J.C., De Dombal, F.T., 1977. Diagnosis of acute abdominal pain in the accident and emergency department. Br J Surg, 64, 250–254.

Wilson, P.D., Horrocks, J.C., Lyndon, P.J., Yeung, C.K., Page, R.E., De Dombal, F.T., 1975 Simplified computer-aided diagnosis of acute abdominal pain. Br Med J, 2, 73–75.

Wilson, S. J., 2017. Data representation for time series data mining: time domain approaches. Wiley Interdisciplinary Reviews: Computational Statistics, WIREs Comput Stat, 9, No.1.

Winsbarg F, Elkin M, May J, 1967. Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. Radiology, 89, 211–215.

Woolf, A.D., Pfleger, B., 2003. Burden of major musculoskeletal conditions. Special Theme – Bone and Joint Decade 2000 –2010, 81, 646-656.

Wright, S., 2009. Optimization Algorithms in Support Vector Machines. University of Wisconsin-Madison Computational Learning Workshop, Chicago, June 2009.

Yeh, C., 2015. Support Vector Machines for classification. Methods and Theory. Available at: http://efavdb.com/svm-classification/.

Yı, J.I., Yu, C., Weıxıng, H., 2016. Processing of Biomedical Weak Signals using Dual-Tree Complex Wavelet Transform. International Journal of Simulation-Systems, Science & Technology, 17, 10.1-10.5.

Zhu, J., Rosset, S., Zou, H., Hastie, T., 2006. Multi-class AdaBoost. Department of Statistics, University of Michigan. Available at http://rob.schapire.net/papers/explaining-adaboost.pdf.

Zupan, J., Gasteiger, J., 1991. Neural networks: A new method for solving chemi cal problems or just a passing phase? Anal. Chim. Acta, 248, 1-30.

Zwarts M. J., Drost, G., Stegeman, D.F., 2000. Recent progress in the diagnostic use of surface EMG for neurological diseases. Journal of Electromyography and Kinesiology, 10, 287-291.

CURRICULUM VITAE

Jean de Dieu UWISENGEYIMANA got his bachelor's degree in Electrical and Electronics Engineering from former Kigali Institute of Science and Technology (KIST) in 2013, Kigali, Rwanda. He then worked as GSM Cell Power Site Engineer for Tigo-Rwanda. He is currently an assistant lecturer at University of Rwanda (UR) and he is completing his Master's degree in Electrical and Electronics Engineering at Çukurova University, Adana, Turkey. His research interests include: Artificial Intelligence, Pattern Recognition, Software Engineering, Biomedical Signal, Image Processing, Computer Vision, Artificial Neural Networks, Machine Learning Algorithms, Biomedical Equipments, Power Systems, Renewable Energy Sources, Electrical Machines and Solar Energy Technologies.

APPENDICES

Appendix 1. Publications Produced Using the data and Results Obtained From This Thesis

Authors	Jean de Dieu Uwisengeyimana and Turgay Ibrikci
Year of Publication	2017
Title	Diagnosing Knee Osteoarthritis Using Artificial Neural Networks and Deep Learning
Journal Volume	2
Series	3
Publisher	Science Publishing Group- Biomedical Statistics and Informatics
Pages	95-102
Intenet Link	http://article.sciencepublishinggroup.com/pdf/10.11648.j.bsi.20170203.13.pdf

: